DATA MINING PROJECT

Master in Data Science and Advanced Analytics

NOVA Information Management School

Universidade Nova de Lisboa

ABCDEats Inc.

Group 17

André Sousa, 20240517

Isabella Costa, 20240685

Jéssica Cristas, 20240488

Tiago Castilho, 20240489

Fall/Spring Semester 2024-2025

TABLE OF CONTENTS

TABLE OF FIGURES

1. ABSTRACT

This project explored customer ordering behavior for ABCDEats Inc. (ABCDE), a fictional food delivery service, by analyzing data collected over three months from three cities. The dataset included transactional details such as customer identities, food types, order frequencies, and spending habits. The primary aim was to identify patterns in customer preferences and segment them into groups with similar behaviors through clustering, providing insights into food trends and marketing strategies.

After preprocessing the data to eliminate missing values and reduce outliers, we performed feature engineering to enhance the dataset with new, relevant features that better capture customer behavior. Following this, the clustering analysis employed K-Means, chosen for its simplicity and efficiency. While alternatives like DBSCAN offered advanced capabilities, K-Means was preferred for its straightforward approach. Metrics such as the silhouette score, coefficient of determination, and Davies-Bouldin score informed the optimal cluster count. Three perspectives guided the clustering process: loyalty and retention, customer behavior, and customer preferences.

These insights were consolidated into a final clustering solution comprising five clusters: consistent moderate spenders, diverse high spenders, new customers, loyal high spenders, and promotion-sensitive customers. Each cluster was analyzed for its unique characteristics, allowing for tailored marketing strategies.

In conclusion, this analysis provided a comprehensive understanding of customer behavior for ABCDEats Inc., enabling the development of targeted marketing strategies. By identifying distinct customer segments and their preferences, ABCDEats can enhance customer retention, boost profitability, and improve overall satisfaction.

2. INTRODUCTION

This project examines customer ordering behavior for ABCDEats Inc. (ABCDE), a fictional food delivery service, using three months of data collected from three cities. The dataset includes transactional details such as customer identities, food preferences, order frequencies, and spending patterns.

Customer segmentation is a widely used approach in data mining and a powerful tool for developing effective marketing strategies. This project focuses on cleaning and preprocessing the collected data to create a robust foundation for segmentation analysis. By applying clustering algorithms, we aim to group customers based on shared characteristics. When analyzed from various business perspectives, these clusters can provide valuable insights into the customer base, enabling the development of targeted strategies to enhance engagement and drive growth.

3. PRE-PROCESSING

3.1. Duplicate Values

In our search for duplicates, we found 26 rows where the customer ID and all other features values are identical. These duplicate rows were dropped from the dataset to avoid redundancies.

3.2. Missing Values

In our search for missing values, we identified a total of three features with missing data: "HR_0", "first_order", "customer_age" and "customer_region".

For "HR_0", we observed that all customers with differences in values between total orders by days and total orders by hours have missing values in the column, which likely represents orders placed during that hour. We used this knowledge to fill in the missing values. The imputation successfully preserved the original proportions, making the chosen method reliable (Figure 1).

For the "first_order" feature, we dropped the rows with missing values since they represented only 0.33% of the dataset. Also, the measures of central tendency were not suitable for accurately imputing values (Figure 2).

For "customer_age", we used the median to fill in the missing values. This decision was made based on the data distribution within this feature, making the median a suitable method for imputation (Figure 1).

For "customer_region," we adopted a more complex approach. Given that our dataset includes three principal regions, characterized by distinct spending patterns on specific types of cuisine, we used this information to address missing ZIP code values. Specifically, we identified the cuisine category with the highest spending for each customer and leveraged the relationship between cuisine preferences and regions to estimate the missing ZIP codes based on the maximum cuisine expenditure of the customers (Figure 4 and Figure 5).

3.3. Categorical Features

To handle the categorical features, we used encoding methods to convert them into numerical features. The “payment_method” and “last_promo” were encoded using the one-hot encoding method, making it easier to further analyze clusters. Although “customer_region” is another categorical feature, there is no need to encode. This feature will only be used for cluster analysis, and one-hot encoding isn’t viable due to the large number of unique values on it.

3.4. Feature Engineering

New features were created to reduce dataset dimensionality and to provide more insights into the customers' behavior, preferences and spending habits.

Total Spending - This feature represents how much money has been spent by each customer on all the dishes they ordered in the three months of study.

Average Spending - The Average Spending that each customer spends on each meal is calculated by Total Spending divided by the Total Orders.

CUI Count - This feature counts the number of different cuisines each customer has ordered from.

Days between orders - This feature calculates the number of days between the first and last order, helping us identify customers who have been ordering over a longer period.

Week and Weekend Orders – This feature counts the number of orders placed on weekdays (Week Orders) and the number of orders placed on weekends (Weekend Orders) by each customer.

3.5. Outlier Removal

We removed outliers by analyzing the boxplots of each metric feature and making precise cuts at points of discontinuity among the outliers. The Interquartile range (IQR) method couldn’t be used for outlier removal, as many features had the first and third quartile equal to zero, which would have resulted in an almost complete removal of the dataset.

Outliers in Cuisine Types, Mealtime, Week Orders, Weekend Orders, Total Spending, Total Orders and Average Spending were handled, as these features were important for clustering. Later, the removed outliers were reintroduced into the clusters.

We ensured that no more than 5% of the data was removed during the outlier's removal, preserving most of the dataset. With this, only 0.90% of the data was removed, which is below the 5% threshold.

3.6. Scaling Data

We applied the Standard Scaling method, specifically the z-score normalization, to our dataset. We chose this approach because it allows outliers, particularly in cuisines, to have dispersed values instead of being compressed between 0 and 1 (as in the case of Min-Max scaling) for clustering.

4. FEATURE SELECTION

We used a Correlation Matrix (Figure 6) to examine the Pearson correlation between the metric features. Features with low correlation were excluded from clustering. Among the strongly correlated features, we retained only a few to reduce the feature space for clustering. Using the remaining features, we applied a soft recursive elimination to retain as many features as possible without losing significant information for our clusters.

5. CLUSTERING

During the selection of the cluster algorithm, we consider the balance between complexity, computational efficiency and sensitivity to high dimensional data. After assessing the good and bad points we opted for K-Means due to its simplicity in implementation. While algorithms like DBSCAN and Self-Organizing Maps offer advanced capabilities and rely heavily on hyperparameter tuning and can be computationally expensive.

Despite its advantages, we acknowledge K-Means tendency to perform optimally with data exhibiting a globular structure and dependence on the initialization.

The main requirement for K-Means is to define the number of clusters before initialization. To make this decision, the measurements coefficient of determination (R2), silhouette score and Davies-Bouldin score were used.

5.1. Clustering Perspectives

For clustering perspectives, we decided to apply segmentation within the business context in a way that would help us in our end goal, providing meaningful customer segmentation that can be translated into marketable strategies ultimately resulting in profit for the company.

In this context, three perspectives were considered: loyalty and retention, customer behavior, and customer preference.

5.1.1. Loyalty and Retention

The use of features related to recency and frequency was applied to create this perspective, using "first_order", "last_order", "Days between Orders", "Total Orders", "Week Orders", "Weekend Orders" to try and identify loyal customers during the period of the data and those who buy less often or had a fall out during the three months of data collection (Figure 7).

No visible patterns were observed for the customer_region in this perspective, as they follow the same trends previously detected in the EDA, reflecting the most frequent regions where they live. Additionally, the "customer_age" does not provide relevant information to identify trends as well.

In this perspective, considering the scores, a solution of 3 clusters was the best one (Table 1).

5.1.1.1. Cluster 0: New and Good Spenders

Composed of customers that started ordering relatively late in the data collection period, they can be characterized as new customers. Due to their recency in first orders, they have yet to formulate preference in cuisine types, reflecting ultimately on a low number of total orders but not out of the ordinary when considering the overall ordering period (Figure 8).

This cluster is one of the biggest ones, accommodating 8732 customers (Figure 9). This group has the potential to be high spenders in the future, considering the average spending (Figure 10) for a relatively low number of total orders, if they are retained in the long run.

5.1.1.2. Cluster 1: Consistent Neutral Spenders

The cluster number one is profiled by the biggest number of people of all clusters (Figure 9), they ordered in all the experiment time starting early and continuing until the end of the experiment, being the most regular clients, but without a high number of orders because the total orders and week orders are both very low considering some of the other clusters. While also showing the second lowest average spending of all the clusters (Figure 11).

In terms of total spending (Figure 11,Figure 12), they spent a medium value, not the highest but also not the lowest, making them neutral.

Now talking about the types of cuisines, the cluster one has a bigger representation in 2 of the cuisines that were evaluated, the Asian, American and the Other. In the rest of the cuisines the cluster doesn't show much of a relevance (Figure 13,**Erro! A origem da referência não foi encontrada.**).

5.1.1.3. Cluster 2: Early Spenders

The customers in this group placed their first orders relatively early but chose not to make additional orders after a short period. Therefore, the “Days Between Orders” metric is very low because there was no major purchase period.

In addition, the “Total Orders” and “Week Orders” indicators are also not notable, showing that the total number of orders made during the study period was scarce. The cluster cardinality is 7988 (Figure 9).

Overall, the total spending was low. These customers have a slightly higher spending in Asian food, but remaining very low compared to the other clusters. Surprisingly, they are the second cluster with a higher “Average Spending” on orders (Figure 10).

5.1.1.4. Cluster 3: Loyal Diverse Spenders

This is the cluster with the least amount of people with only 2307 clients (Figure 9). However, they are the most loyal clients who started spending it right from the beginning and they never stopped making orders (Figure 13). This made them the group of customers who spent the most money (Figure 11) just by being regulars, because they are also the customers who had the least average spending on orders (Figure 10). These customers are also the ones who like to vary more in their orders, making them the ones who spent the most in the American, Asian, Beverages, Chicken Dishes, Chinese, Indian, Italian, Japanese and Other cuisines (Figure 13,Figure 14).

5.1.2. Customer Behavior

For this perspective we tried to analyze emerging patterns related to buying habits related to frequency, time of the week, mealtime and amount spent. To achieve this, we considered initially the features "Total Orders", "Days between Orders", "Week Orders", "Weekend Orders", "CUI Count", "first_order", "last_order", "Average Spending", "Early morning", "Breakfast", "Lunch", "Afternoon Snack", "Dinner" and "Supper" (Figure 16).

No visible patterns were observed for the "customer_region" in this perspective, as they follow the same trends previously detected in the EDA, reflecting the most frequent regions where they live. Additionally, the "customer_age" does not provide relevant information to identify trends as well.

For this perspective, arrived at an optimal solution of 3 clusters (Table 2 – Perspective 2 scoresTable 2).

5.1.2.1. Cluster 0: Late High Spenders

This cluster is composed of customers that started ordering at the end of the data collection so have a small record of orders and did not spend a lot of money (Figure 17), underlying patterns are not visible in terms of cuisines or mealtimes (Figure 18,Figure 19), but when they order they spend a good amount on average (**Erro! A origem da referência não foi encontrada.**), and the highest when compared to other clusters. This cluster is composed of 8745 (Figure 21), being the second biggest cluster.

5.1.2.2. Cluster 1: Steady High Spenders

Cluster number one is profiled by the customers that start ordering first and stop ordering the latest of all the clusters (Figure 17). It has a significant volume of orders and a high level of spending. In terms of weekends and weekdays it shows a high number of both showing just a slightly lower number for weekend orders (**Erro! A origem da referência não foi encontrada.**).

It's also the cluster with the lowest number of customers but with the highest amount of spending of all clusters (**Erro! A origem da referência não foi encontrada.**), showing also a high diversity of cuisines (**Erro! A origem da referência não foi encontrada.**).

5.1.2.3. Cluster 2: Early Light Spenders

The customers in the second cluster placed a low number of orders and it is the cluster that spent the least amount of money (Figure 17). It is characterized by having customers who placed orders early but also stopped ordering just as early. Also, these customers didn't diversify the types of cuisines (**Erro! A origem da referência não foi encontrada.**) they were consuming since they have a low "CUI Count".

The cluster cardinality is 7924 (Figure 21). These customers have a slightly higher spending in Asian food, but remaining very low compared to the other clusters. Surprisingly, they are the second cluster with a higher Average Spending on orders.

Their consumption was higher during the "Breakfast", "Lunch" and "Afternoon Snack" mealtime.

5.1.2.4. Cluster 3: Moderate Mass Spenders

This cluster is the one with the greatest number of people, with 12052 customers. However, they were only the second cluster with the most total spending. The customers of this cluster started near the beginning of the dataset and stopped near the end. They are the second group of customers with the least average spending. They are also the second cluster with the most days between orders.

5.1.3. Customer Preference

In this last perspective we considered customer preference based on the main selling point of ABCDEats, food. Here we aimed to find distinctive patterns related to customer preference based on all the cuisines.

The assessment of the number of clusters resulted in a final selection of 3 (Table 3).

5.1.3.1. Cluster 0: Afternoon Cuisine Enthusiasts

The customers in this cluster spent more in the Café, Italian, Other and Thai cuisines. They also are the ones who order the most in the afternoon. It's one of the two clusters with the least number of customers, only having 1980.

5.1.3.2. Cluster 1:

The cluster number one is profiled by the biggest number of people of all clusters, with a total of 27203 customers, they have an equal spending in terms of the different cuisines with slightly more on American and Asian but have a low amount of spending and a low number of orders when compared to the other clusters.

As for the customer distribution by region, we can see that the regions 2360, 4660 and 8670 are the ones where most of the customers are located.

5.1.3.3. Cluster 2: High-Spending Food Enthusiasts

The customers in the second cluster have very high spending on Asian food, desserts, street food and snacks, above the mean not only for the cuisines spending but also compared to the other clusters. Although their consumption is significantly lower compared to the other cuisines, these customers still consume a small amount of American food.

The cluster cardinality is 2164, which is very low compared to cluster 1. Their most regular mealtimes for ordering are “Early Morning”, “Breakfast”, “Lunch” and “Supper”. It's the cluster with the higher “Average Spending”. Most of them reside in region 8670, and none of them live in region 2490.

6. FINAL SOLUTION

In this section, all the clusters resulting from the different perspectives were merged and further analyzed to achieve an end cluster solution for segmentation. A hierarchical cluster was done to access clusters of proximity, close ones where merged, resulting in a solution of 5 clusters.

The 284 outliers removed (Outlier Removal) were inserted back in the final solution after training a Decision Tree [2] model with the classification of the merged labels as the target variable. Most of the outliers were classified as belonging to cluster 0, corresponding to 62% of the outliers, while the rest were more evenly distributed with cluster 1 accounting for 17% of observation, cluster 2 with 12%, cluster 3 with 9%. No outliers were classified as belonging to cluster 2. No visible alteration on feature average was found when comparing before and after results, indicating that the outliers did not disturb the values of the clusters to which they were added (Figure 40,Figure 41)

6.1. Cluster 0: Consistent Moderate Spenders

This cluster has the largest number of people, with a total of 12320 customers. It ranks third in terms of higher total spending among the clusters and the second with the least average spending per cluster. The customers started ordering early and stopped ordering later in the three-month study period.

Although they don't have a preferred type of cuisine, the cuisines with a more significant spending are the American and Asian. They tend to make more purchases without promotions than with them.

6.1.1. Marketing Strategy

Since this group of customers doesn't care too much about promotions, we should focus on offering value-added services rather than price reductions. Provide faster delivery options, personalized meal recommendations, or exclusive access to premium menu items. These Improvements reinforce the perception of quality, encouraging repeat orders without the need for promotional discounts.

In addition, using engaging content to naturally encourage purchases by inspiring interest in cuisines and dishes. Share compelling stories about trending dishes, chef-curated specialties, or behind-the-scenes insights into how meals are prepared. Leveraging newsletters, blog posts and social media to highlight popular American and Asian dishes that align with their spending habits. This strategy builds trust, motivating customers to order more frequently.

6.2. Cluster 1: Diverse High Customers

As for cluster one, the number of consumers is 1900, being the second smallest cluster, having customers that ordered from the beginning of the study period until the end.

Relative to the cuisines consumed by these customers, Italian and American are the most consumed ones, and cuisines like Desserts, Asian and Street food/Snacks are all close to the average values. It's the cluster that has a higher consumption diversity, as the "CUI Count" is the biggest.

The “Total Orders” is very high, with a “Total Spending” relatively low when compared to the number of orders. This characteristic makes this cluster the one with the smallest “Average Spending”, while being the second cluster with the biggest “Total Spending”.

As for the time of consumption, it shows a major consumption on weekdays being the biggest consumers on weekdays, while nevertheless there is still a big, consumption on weekends.

Now talking about the mealtimes, the orders are normally placed for “Breakfast”, “Lunch”, “Afternoon Snack” and “Dinner”. Most customers are from regions 4660 and 2360, with no customers residing in region 8370.

6.2.1. Marketing Strategies

This cluster shows peak consumption on weekdays, so we should incentivize repeat orders through loyalty programs specific for these days. Offering loyalty points or exclusive discounts for consecutive weekday purchases. One example could be providing a *Weekday Streak Bonus* for customers who order multiple times during the weekdays to boost engagement and spending.

In addition, with high order frequency but low spending, a tiered discount system can encourage larger purchases. For example, offering progressively higher discounts or rewards as the “Total Spending” within a single order or a specific time frame increase. This approach leverages their existing behavior of frequent ordering while motivating them to spend more per transaction.

Finally, given their regional concentration in 4660 and 2360, targeted promotions during peak mealtimes (lunch, afternoon snacks, and dinner) can be highly effective. Highlight Italian and American dishes, which are the top preferences, and bundle these cuisines with complementary items like desserts or beverages.

6.3. Cluster 2: New Customers

In this cluster, we have potential new customers based on the interval between their first and last order, as both were placed later in the study period. Their number of orders is also small, but they cannot be dismissed as poor spenders, as their spending on average is not insignificant given the number of orders they place. Due to the lack of records during the data collection period, this cluster is not specifically characterized by their preference for certain days to order or times they prefer to place orders. However, they tend to match the average at “Supper” and “Dinner” times.

No distinct patterns were detected in food preferences. This cluster is composed of 8732 customers, mainly distributed through the most popular regions (8670, 4460, 2360). For now, they stand out for marketing purposes but have the potential to become high spenders in the long run if they stay loyal and keep ordering.

6.3.1. Marketing Strategies

Firstly, to make a strong first impression, offering a warm onboarding experience with incentives like welcome discounts, free delivery for their next order, or introductory deals on popular cuisines. These initial offers can encourage repeat purchases and cultivate a sense of value, helping to establish our platform as their go-to platform for food delivery.

In addition, introducing a points-based rewards system that incentivizes continued engagement. Customers can earn points for every order, which can later be redeemed for discounts, exclusive menu items, or other perks.

6.4. Cluster 3: Loyal high spenders

This cluster is the smallest one, composed of only 407 customers that relate to ordering throughout the whole period of data collection, and considering their high number of orders and total spending we can qualify them as loyal high spenders since, on average, they are also the customers that spend the most. They order more during the weekdays and mainly for early morning, breakfast and supper. Here we can also see a clear preference for Asian food, street food and snack, deserts, American and Japanese. In terms of region, 80.6% of the customers are situated in customer region 8670.

6.4.1. Marketing Strategies

Cluster 3 represents a highly valuable customer group due to their consistent ordering patterns and spending habits. Naturally it's in the company's best interest to keep them as the loyal customers they currently are.

An exclusive VIP Program can be implemented to reward these customers and strengthen their connection with ABCDEats Inc. Incentives rewards to those who have high spending like priority delivery and cuisine discounts can be applied in this case.

Furthermore, their concentration on customer region 8670 can be explored when offering discounts associated with their favorite. These strategies aim to keep this cluster's loyalty and preference in the long run.

6.5. Cluster 4: Promotion-Sensitive

It's the cluster with the least amount of total spending while still having 7988 customers. However, they have an average spending per cluster above average and don't have a preferred cuisine type.

They are customers who ordered early and the only ones who stopped very early as well and have the least number of total orders. They don't show significant preference towards the hours in which they make purchases. It's one of the clusters who show a bigger number of orders when there's a promotion.

6.5.1. Marketing Strategies

To target this group of customers in marketing strategies, ABCDEats Inc. can target their low number of total orders and early drop-out behavior, while focusing on their main characteristic as buyers.

To do this, loyalty rewards based on promotion can be applied to increase customer retention. Additionally, offering "We Miss You" offers with time-sensitive discounts to encourage streak buying, can be good strategies to increase spending and frequency for this cluster.

7. CONCLUSION

In conclusion, our analysis of customer ordering behavior for ABCDEats Inc. has provided valuable insights into the preferences and trends of customers. By analyzing key features such as order frequency, food category popularity, and the tendency to order from chain versus non-chain restaurants, we identified distinct patterns in customer behavior. We then managed to group those customers into clusters according to their similar preferences and were able to provide specific marketing strategies for each of them.

8. BIBLIOGRAPHY

[1] Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics, 9*(8), 1295. https://doi.org/10.3390/electronics9081295

[2] Scikit-learn developers. (n.d.). *Tree-based models — scikit-learn 1.5.0 documentation*. Retrieved January 6, 2025, from https://scikit-learn.org/1.5/modules/tree.html

APPENDIX

Table 1 – Perspective 1 scores

Number of Clusters	R^2	Silhouette Score	Davies-Bouldin Index
2	0.39	0.37	1.06
3	0.56	0.38	0.87
4	0.72	0.42	0.75
5	0.77	0.40	0.82
6	0.80	0.36	0.94
7	0.82	0.35	0.92

Table 2 – Perspective 2 scores

Number of Clusters	R^2	Silhouette Score	Davies-Bouldin Index
2	0.36	0.40	1.11
3	0.52	0.30	1.18
4	0.61	0.32	1.00

5	0.67	0.30	1.15
6	0.69	0.28	1.33
7	0.71	0.25	1.40

Table 3 – Perspective 3 scores

Number of Clusters	R^2	Silhouette Score	Davies-Bouldin Index
2	0.12	0.58	1.62
3	0.23	0.61	1.77
4	0.32	0.58	1.54
5	0.40	0.58	1.34
6	0.48	0.57	1.15
7	0.44	0.56	0.95

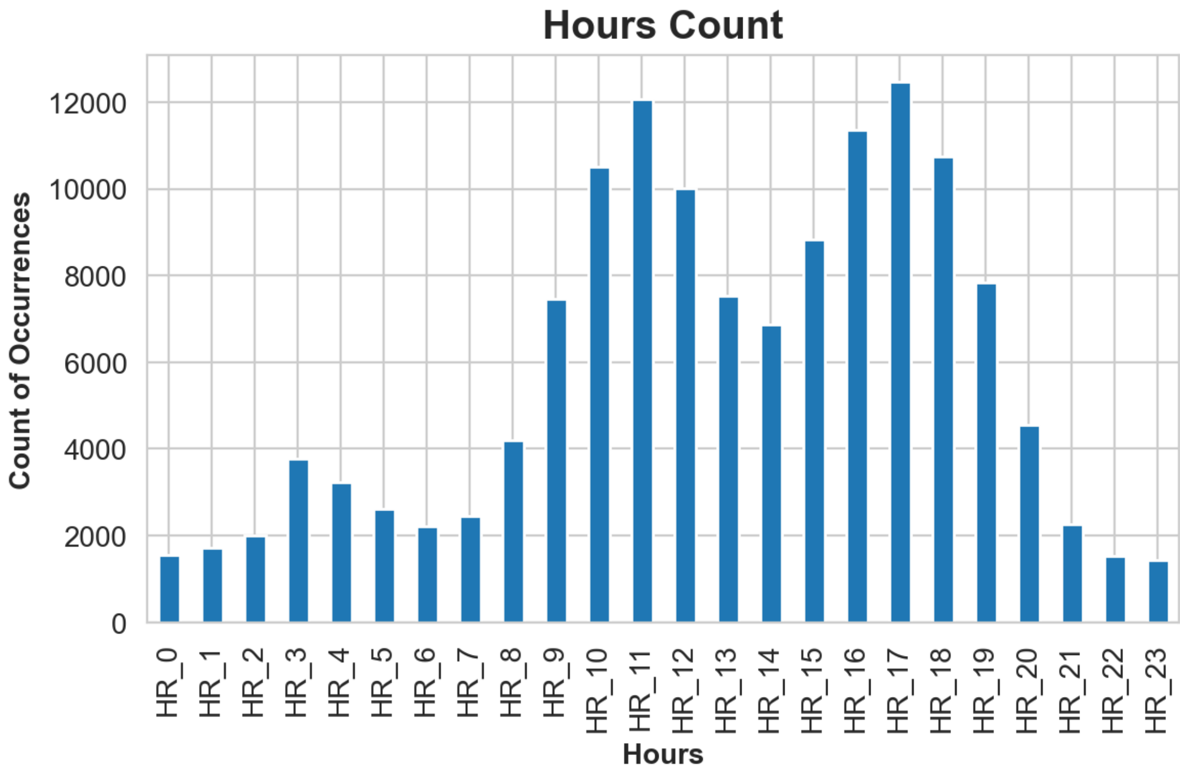

Figure 1 - Frequency for each hour feature after filling the missing values in ”HR_0”

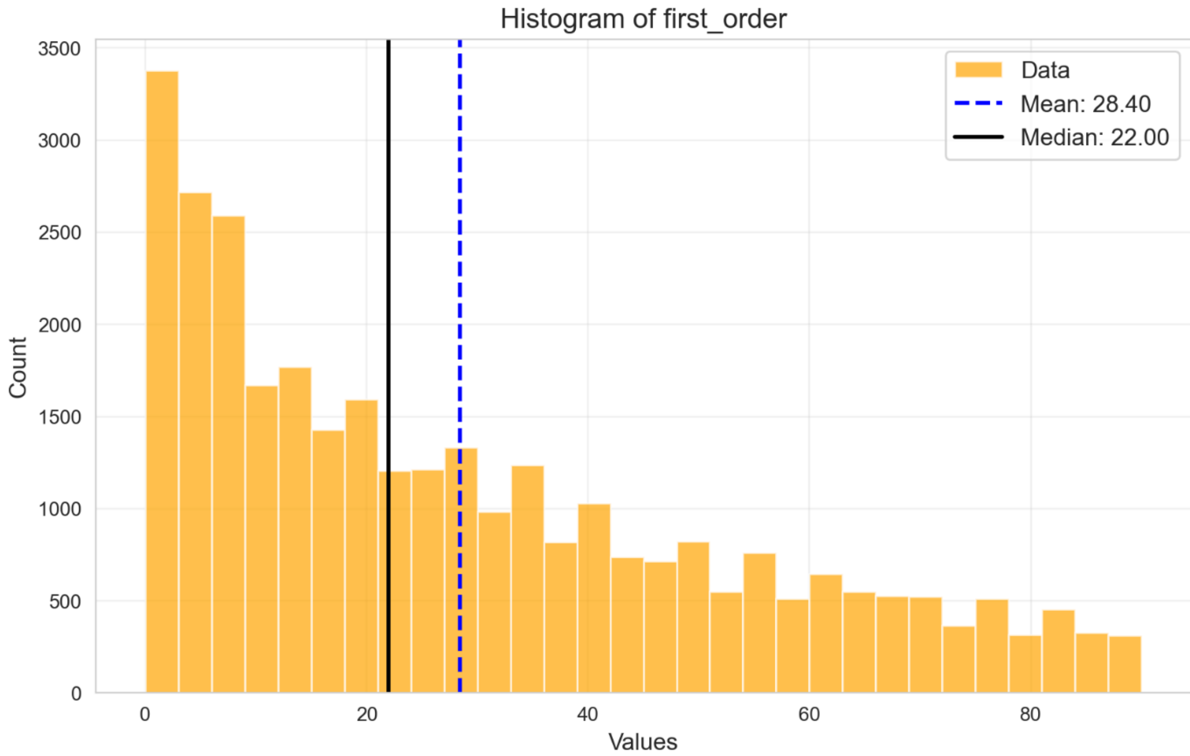

Figure 2 - Frequency and measure of central tendency for ”first_order”

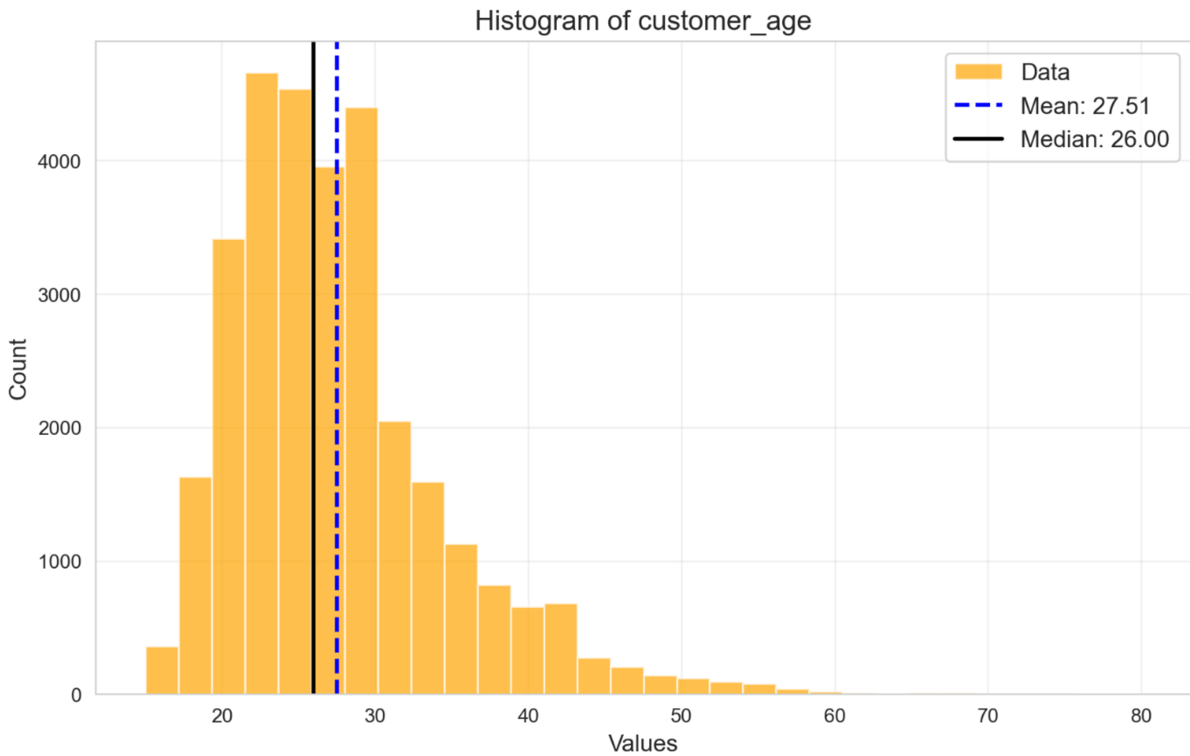

Figure 3 - Frequency and measure of central tendency for ”customer_age”

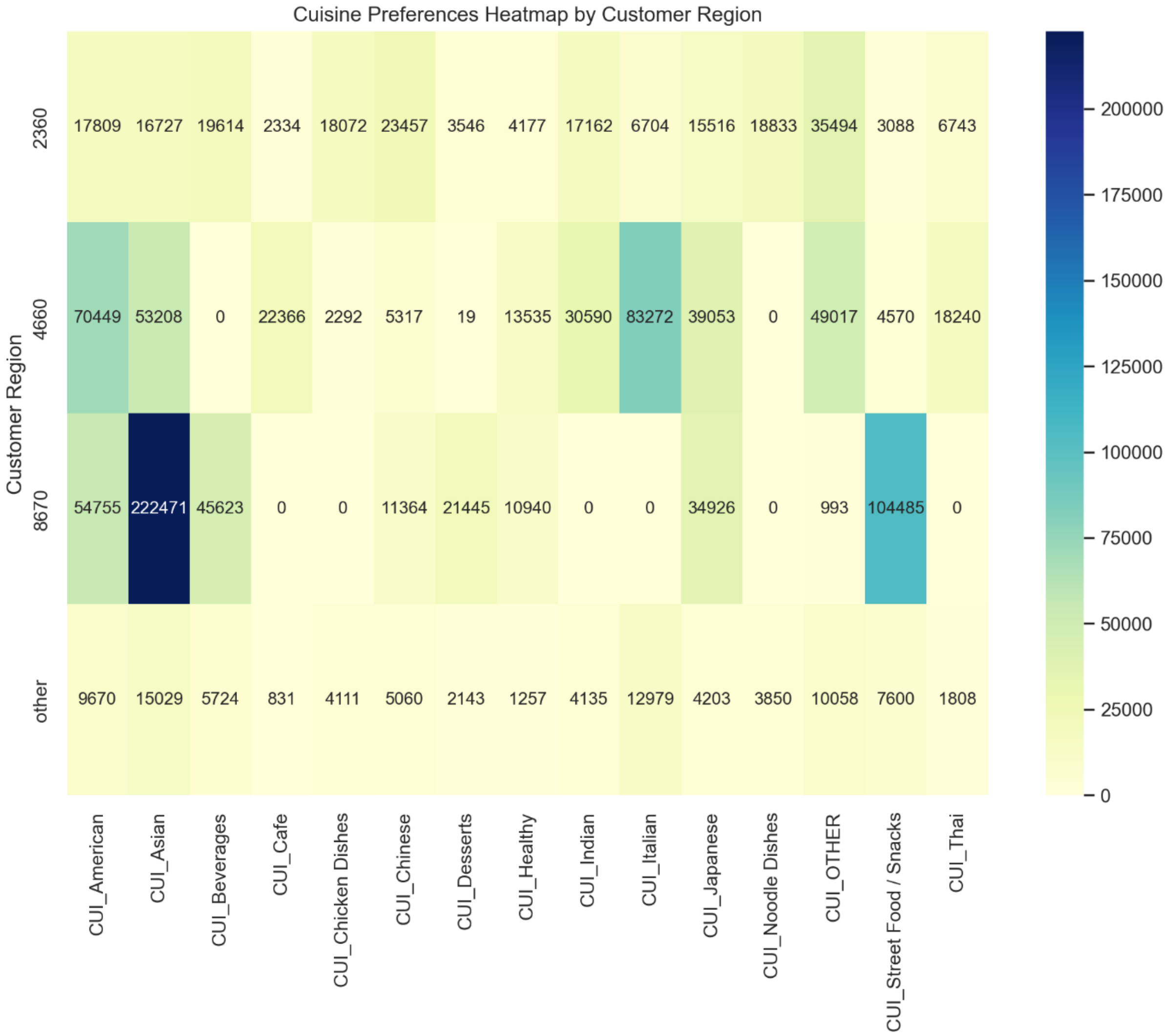

Figure 4 - Relation between the ”customer_region” and the CUI types most spended

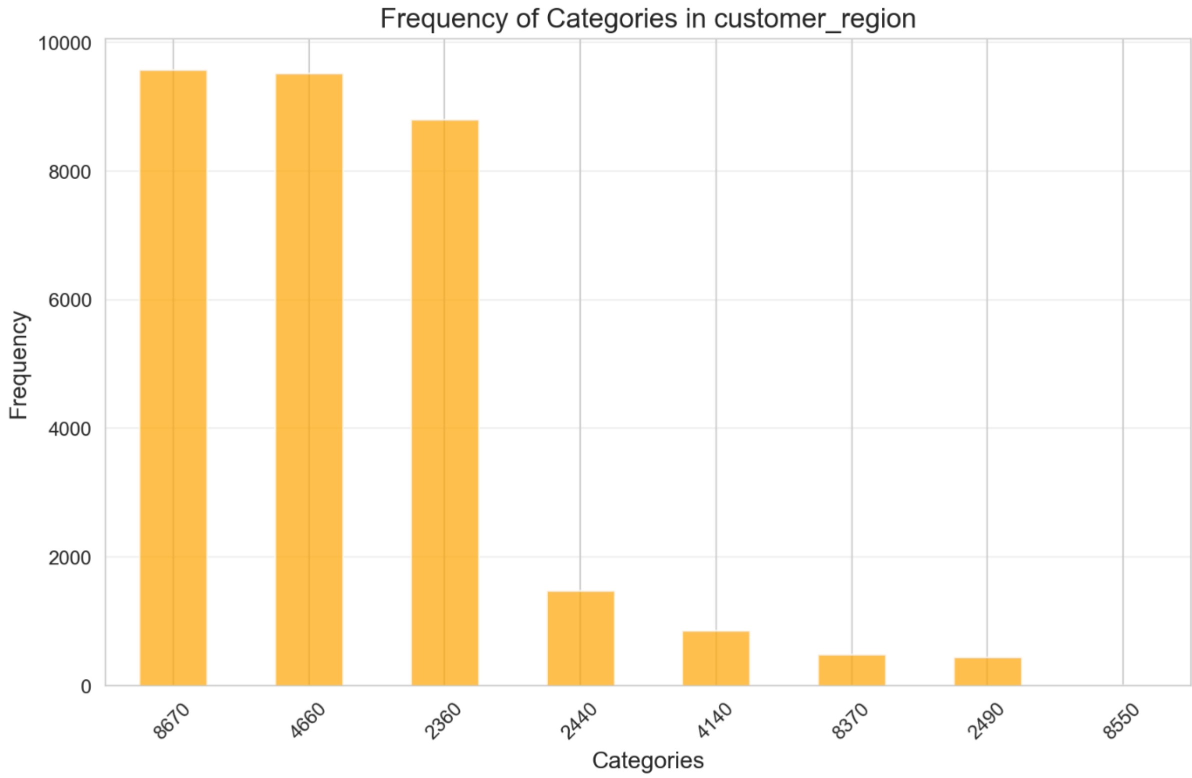

Figure 5 - Frequency “customer_region”

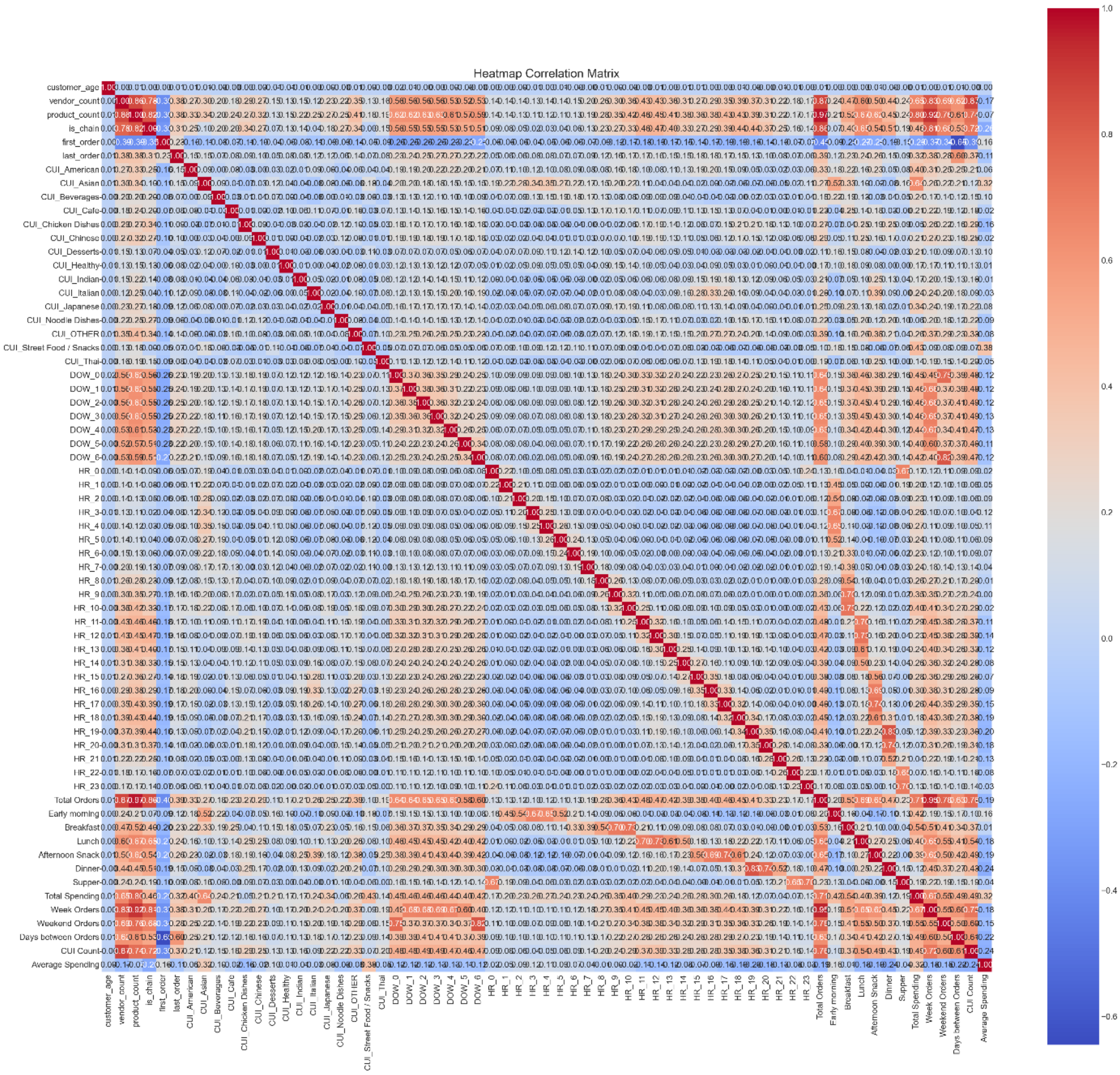

Figure 6 – Correlation Matrix

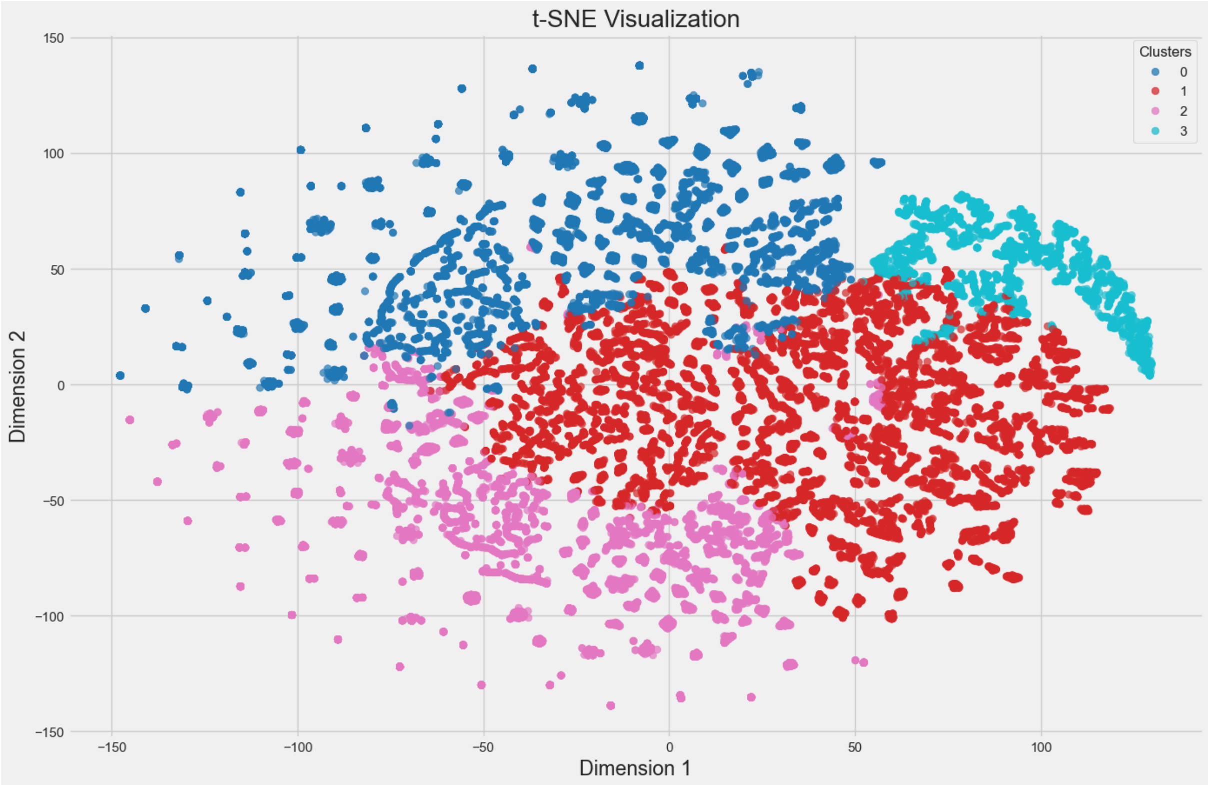

Figure 7 - t-SNE perspective 1

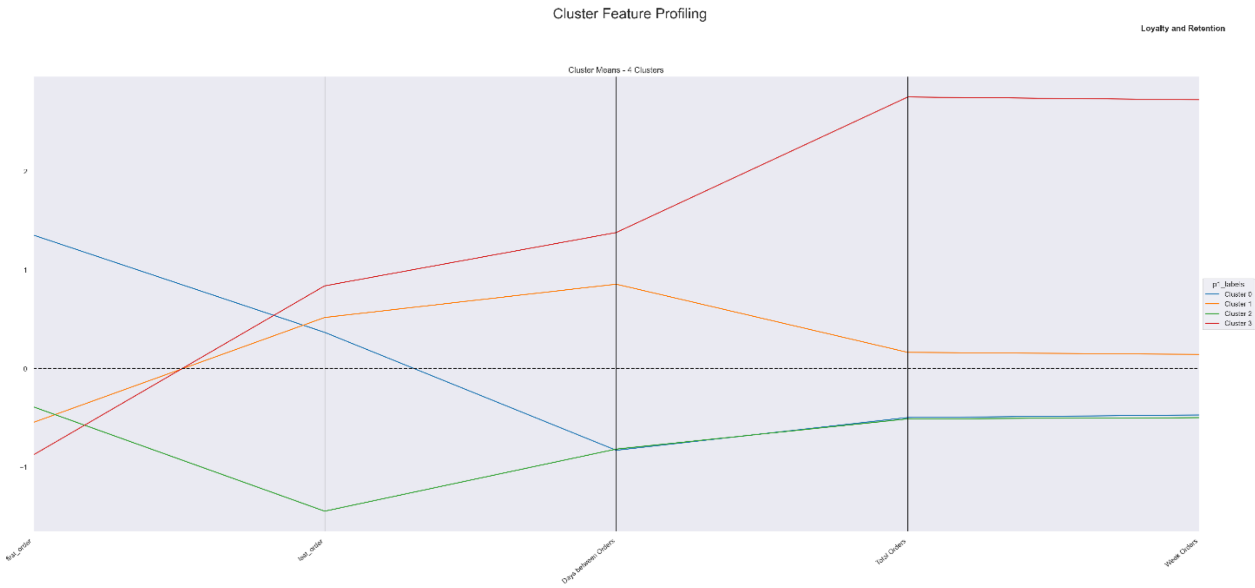

Figure 8 – Cluster feature profiling: perspective 1

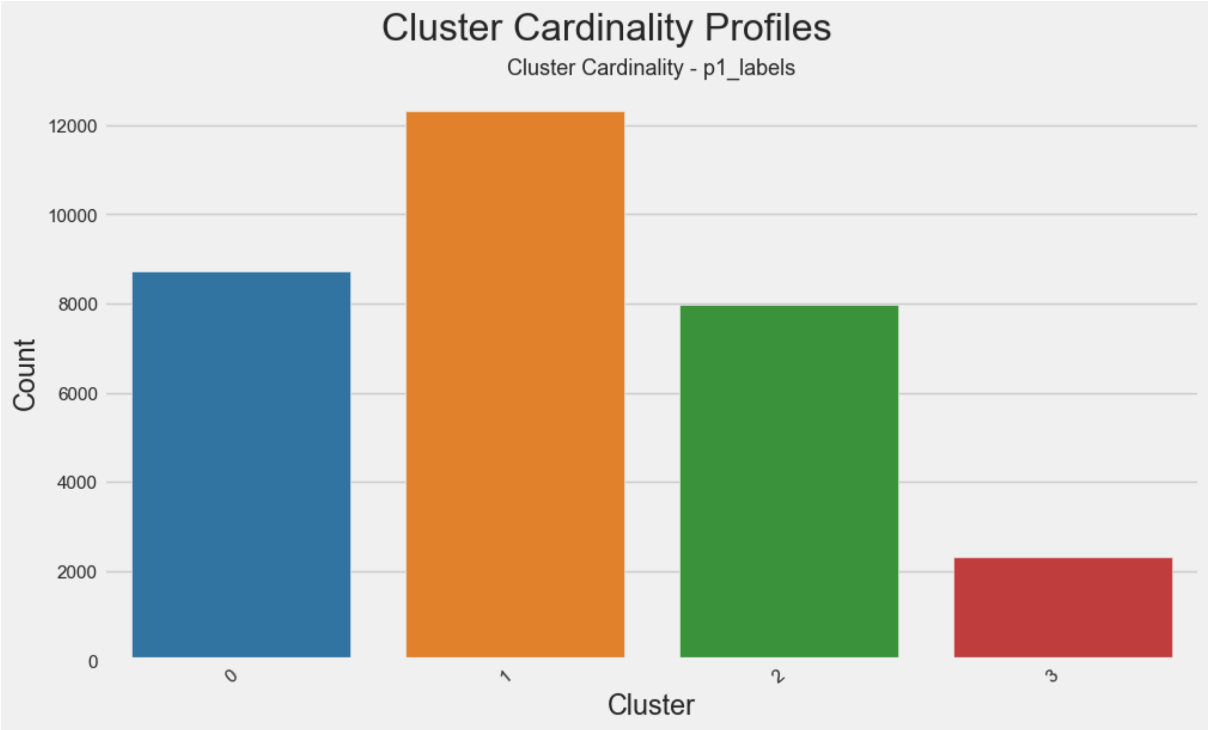

Figure 9 - Cardinality perspective 1

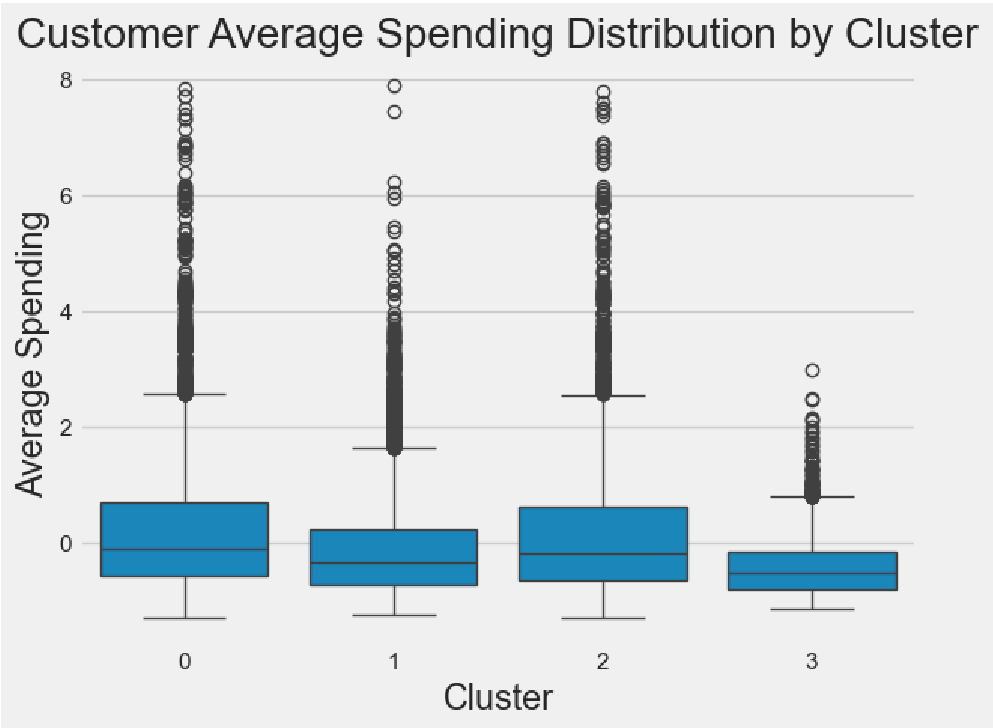

Figure 10 - Average spending boxplots for perspective 1

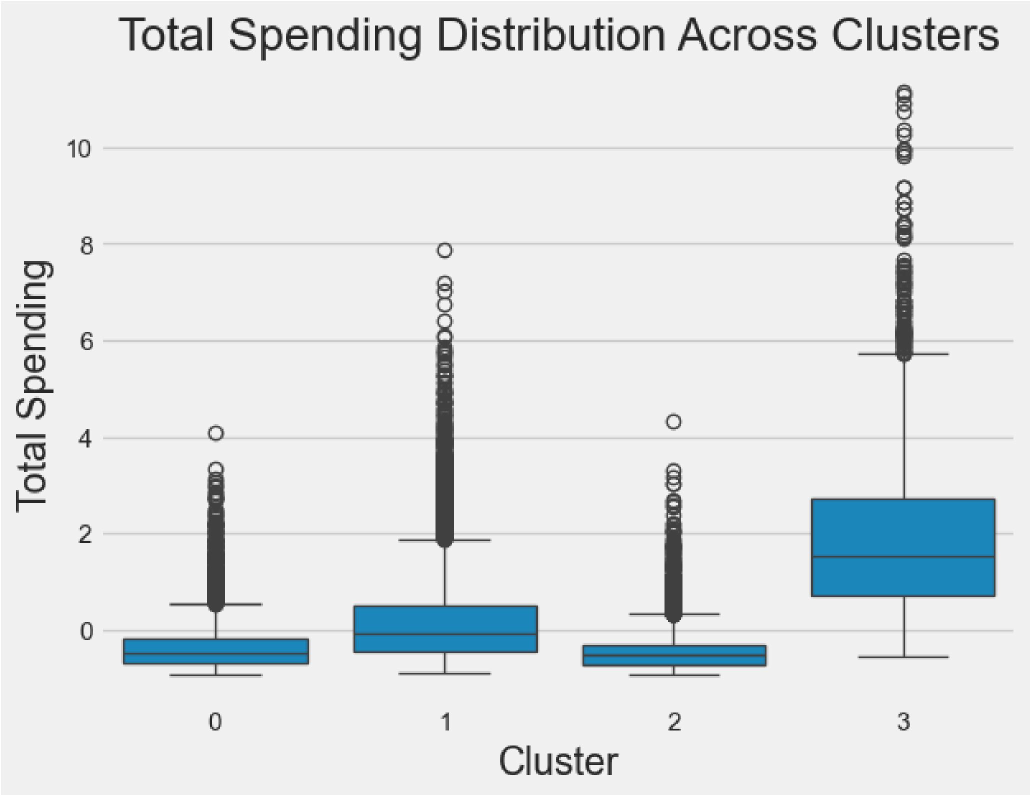

Figure 11 - Total spending boxplot for perspective 1

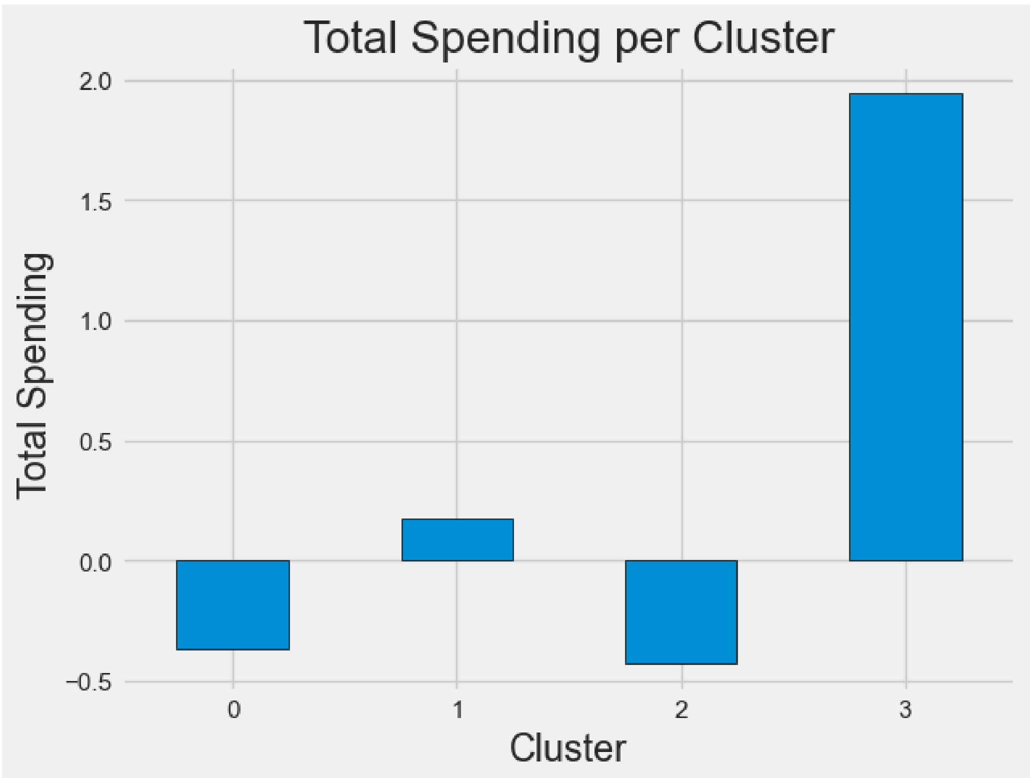

Figure 12 - Total spending bar plot for perspective 1

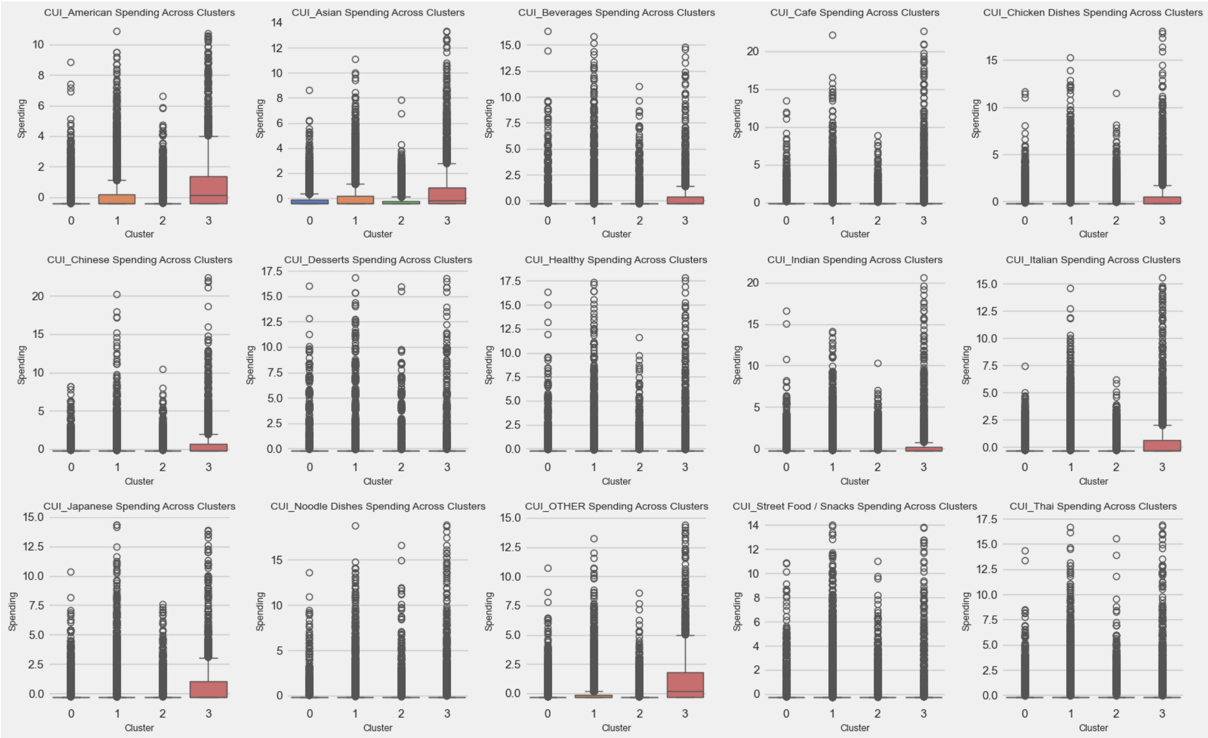

Figure 13 - Cuisine types boxplot for perspective 1

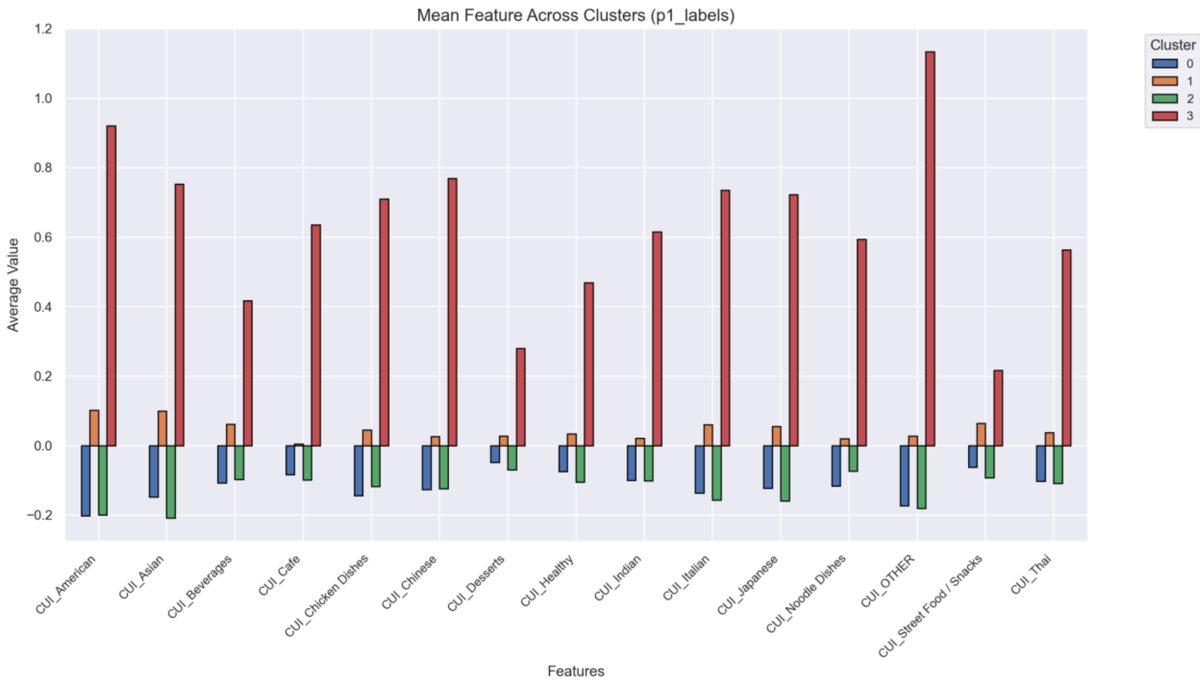

Figure 14 - Cuisine type bar plots for perspective 1

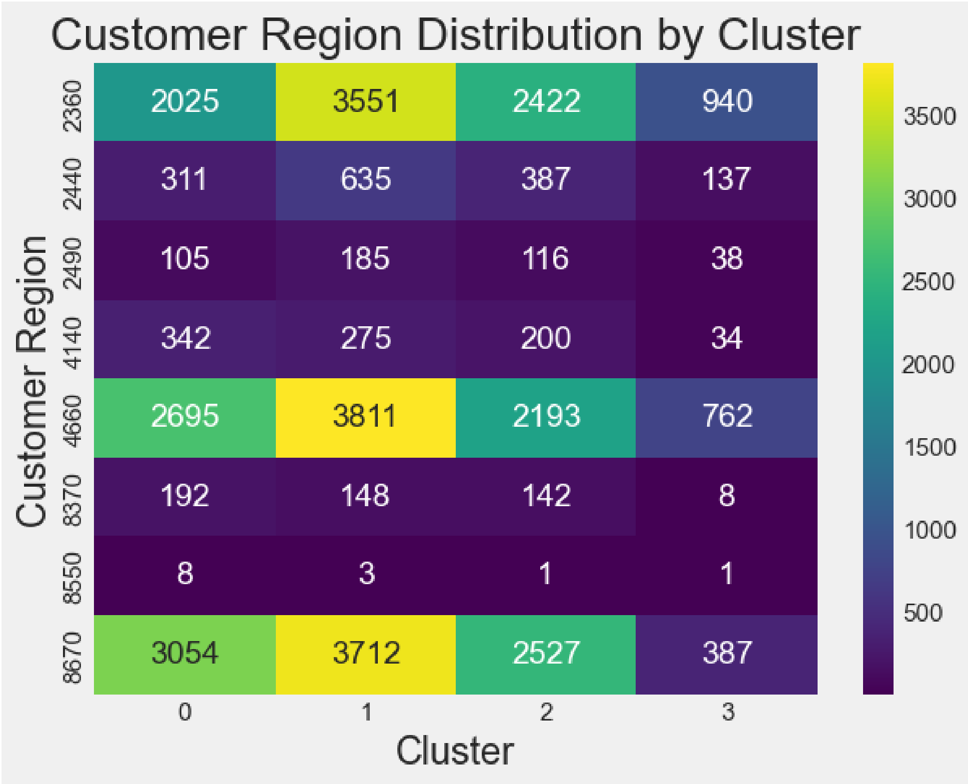

Figure 15 – Customer region distribution for perspective 1

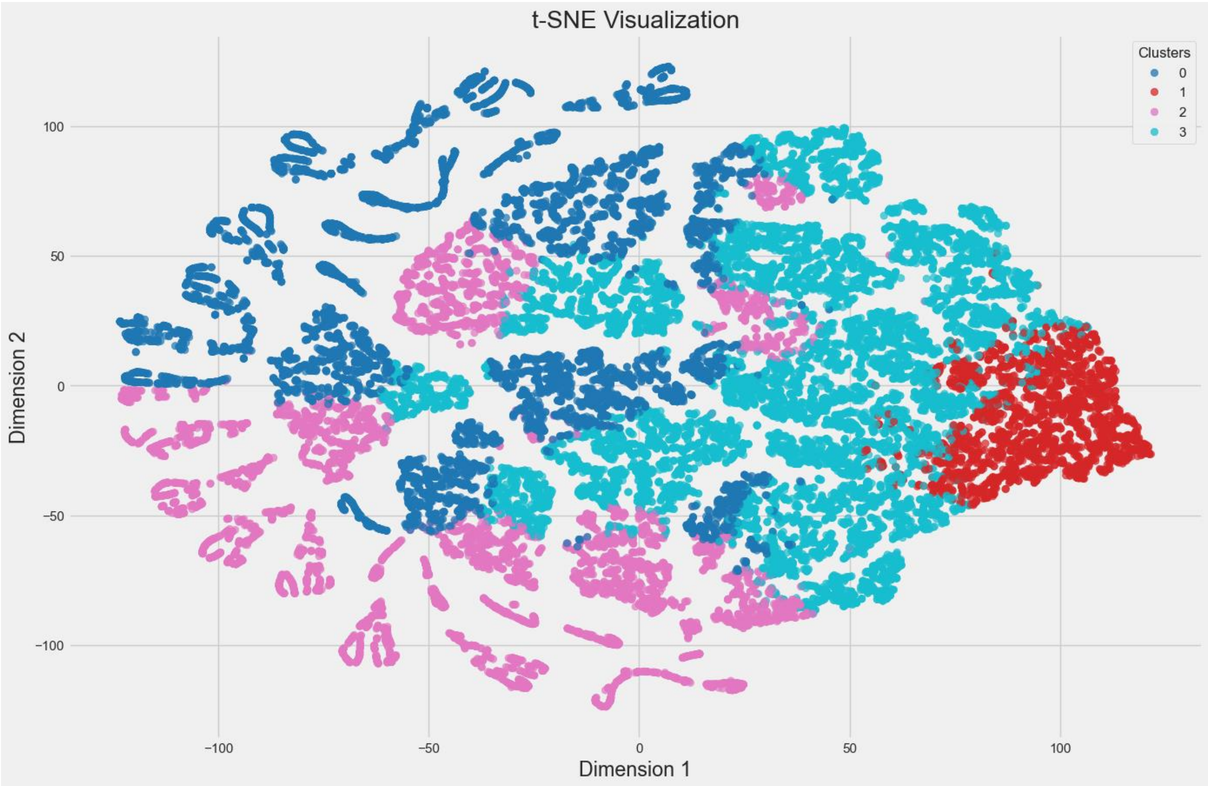

Figure 16 – t-SNE Perspective 2

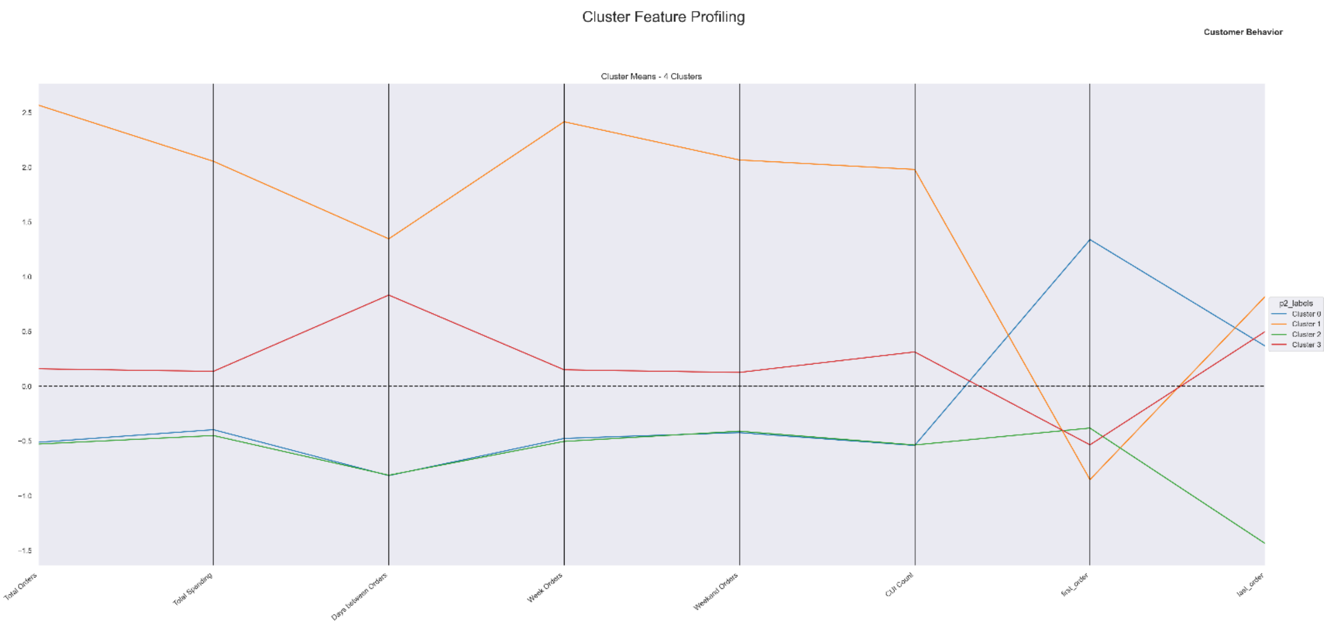

Figure 17 – Cluster feature profiling for perspective 2

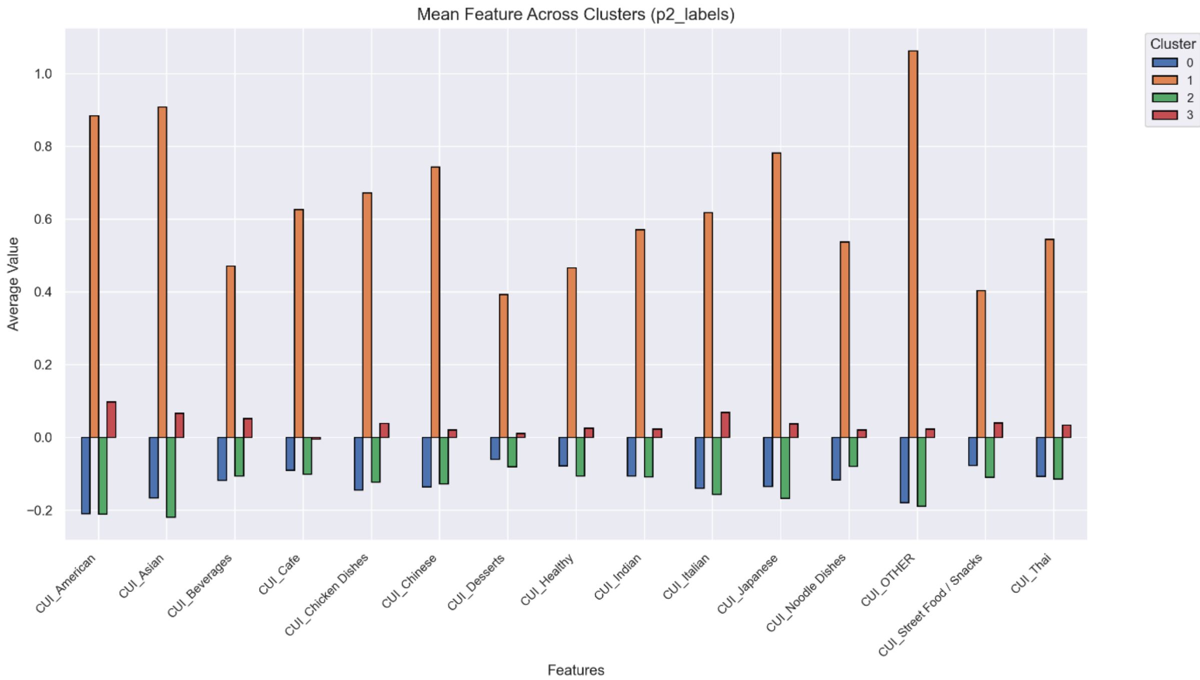

Figure 18 - Cuisine type bar plot for perspective 2

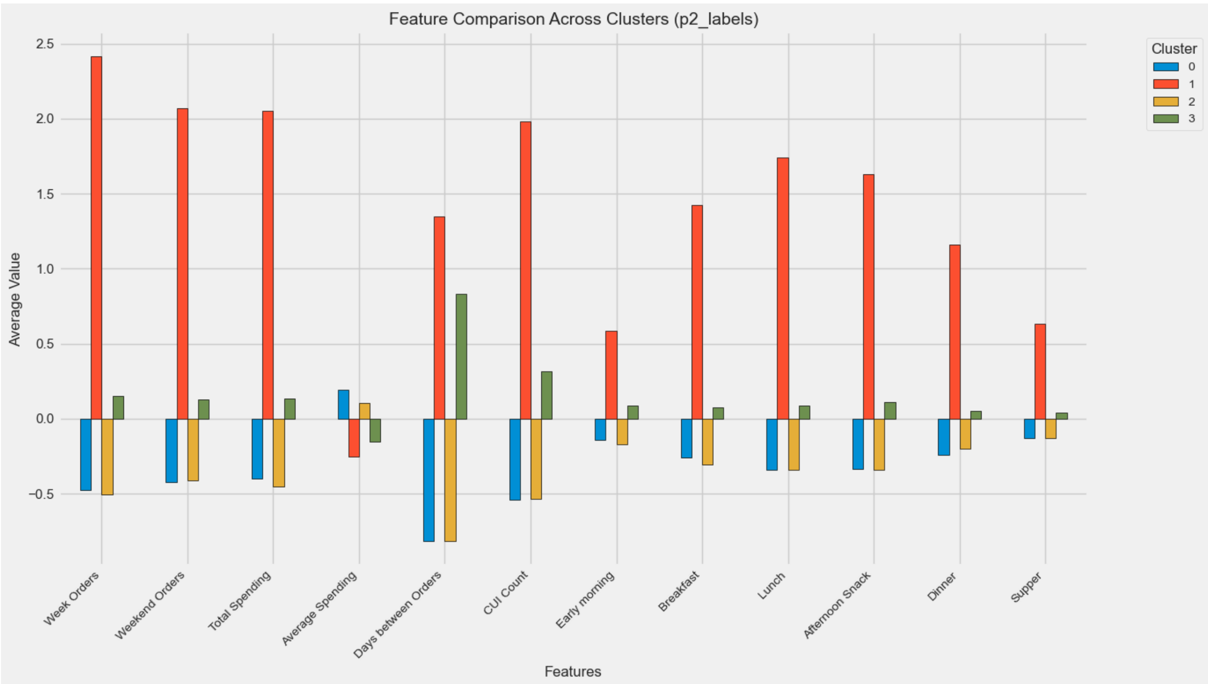

Figure 19 - Unused numeric features bar plot for perspective 2

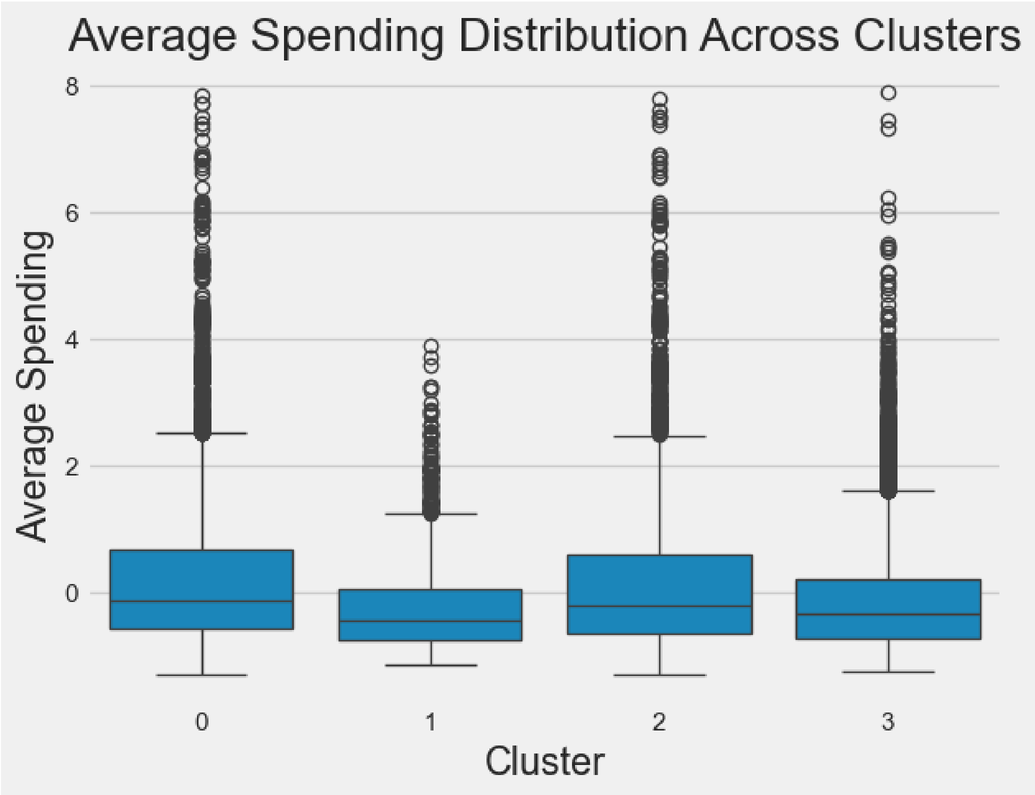

Figure 20 - Average spending boxplots for perspective 2

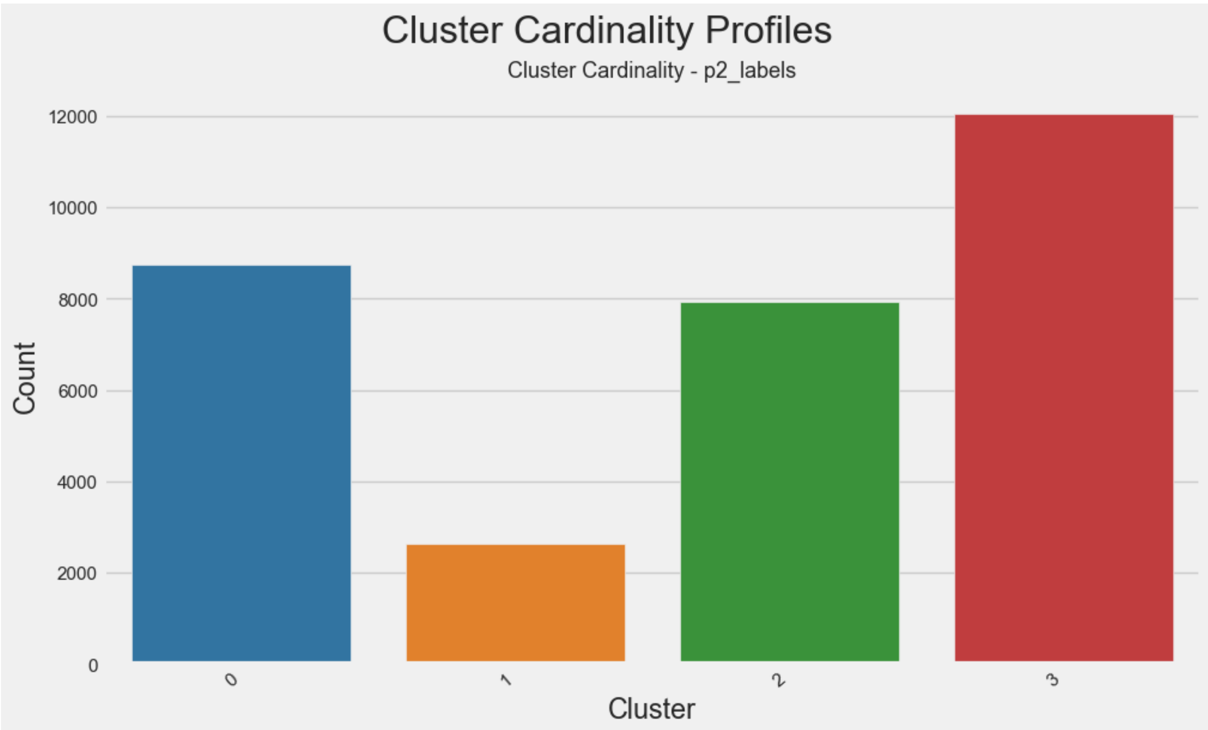

Figure 21 – Cluster cardinality for perspective 2

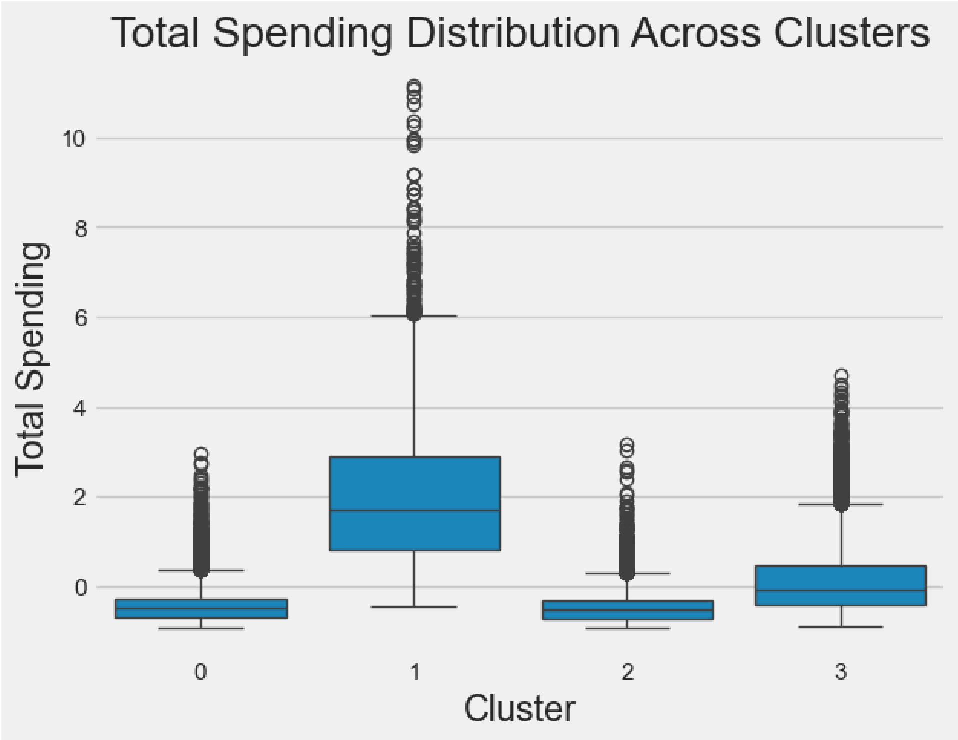

Figure 22 - Total spending boxplot for perspective 2

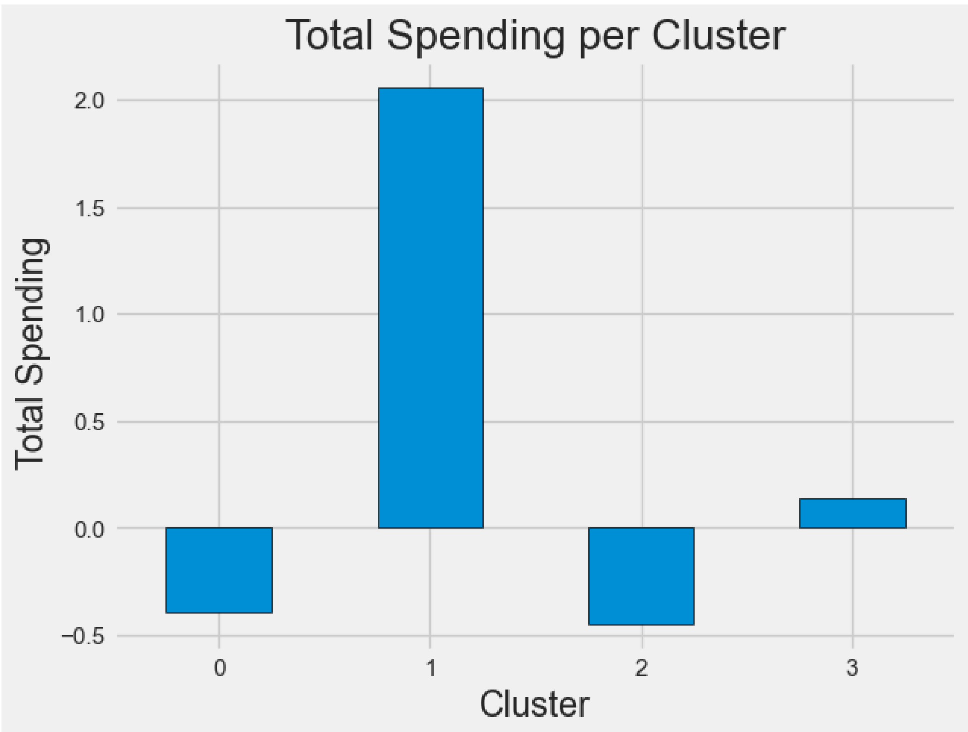

Figure 23 – Total spender bar plot for perspective 2

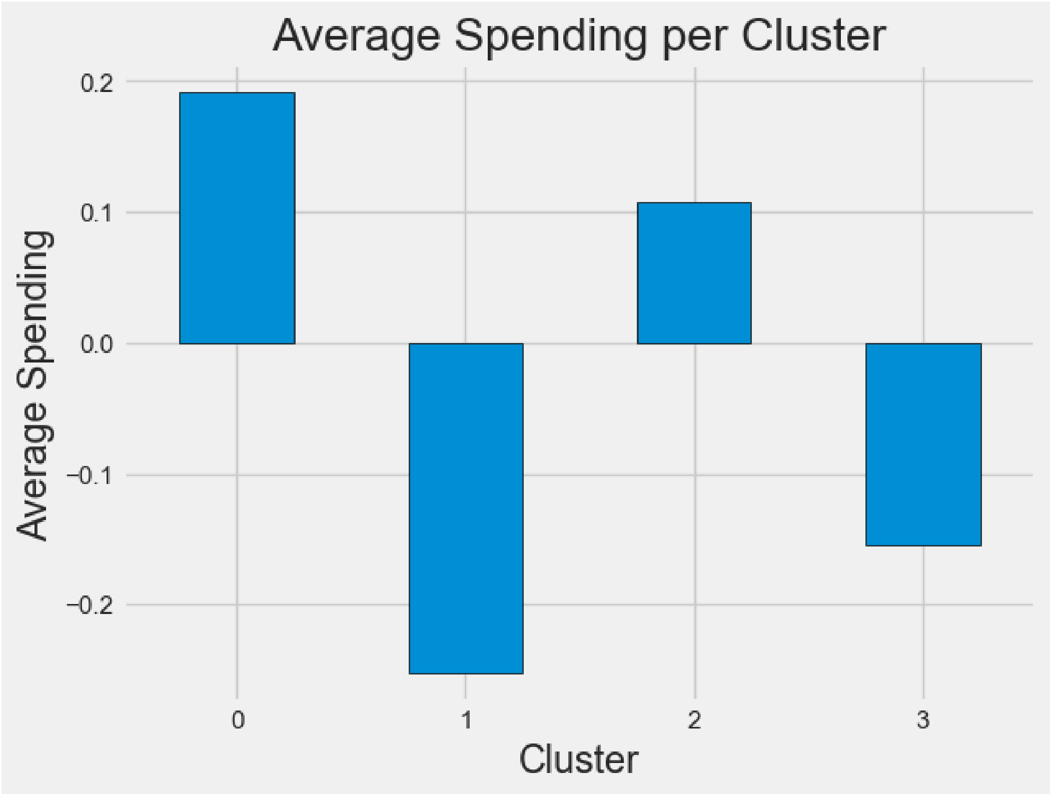

Figure 24 – Average spending bar plot for perspective 2

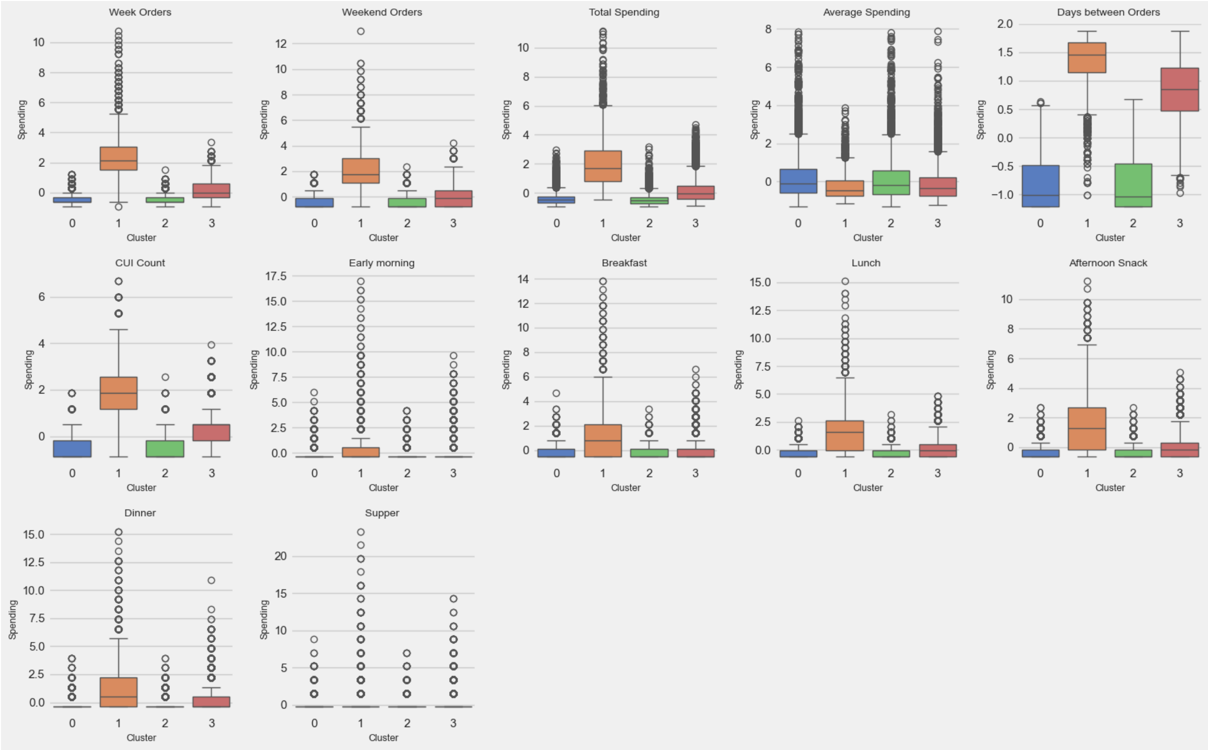

Figure 25 - Unused numeric features boxplots for perspective 2

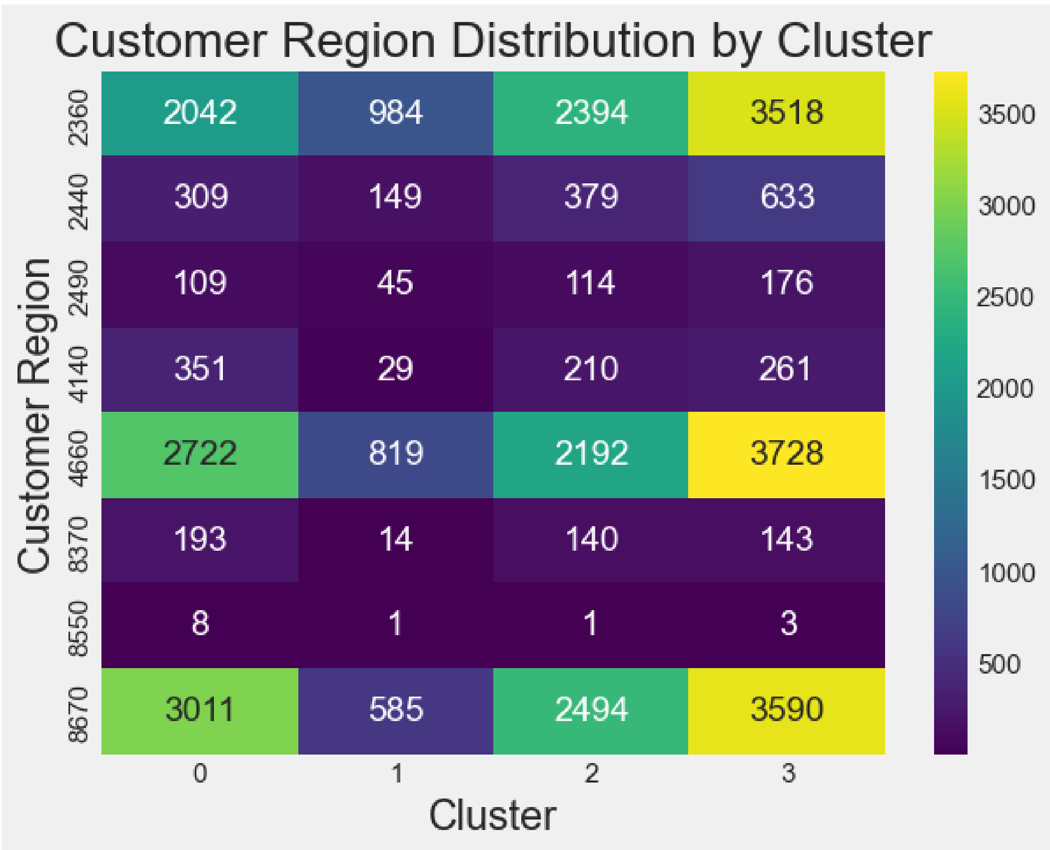

Figure 26 – Customer region by cluster for perspective 2

PERSPECTIVE 3

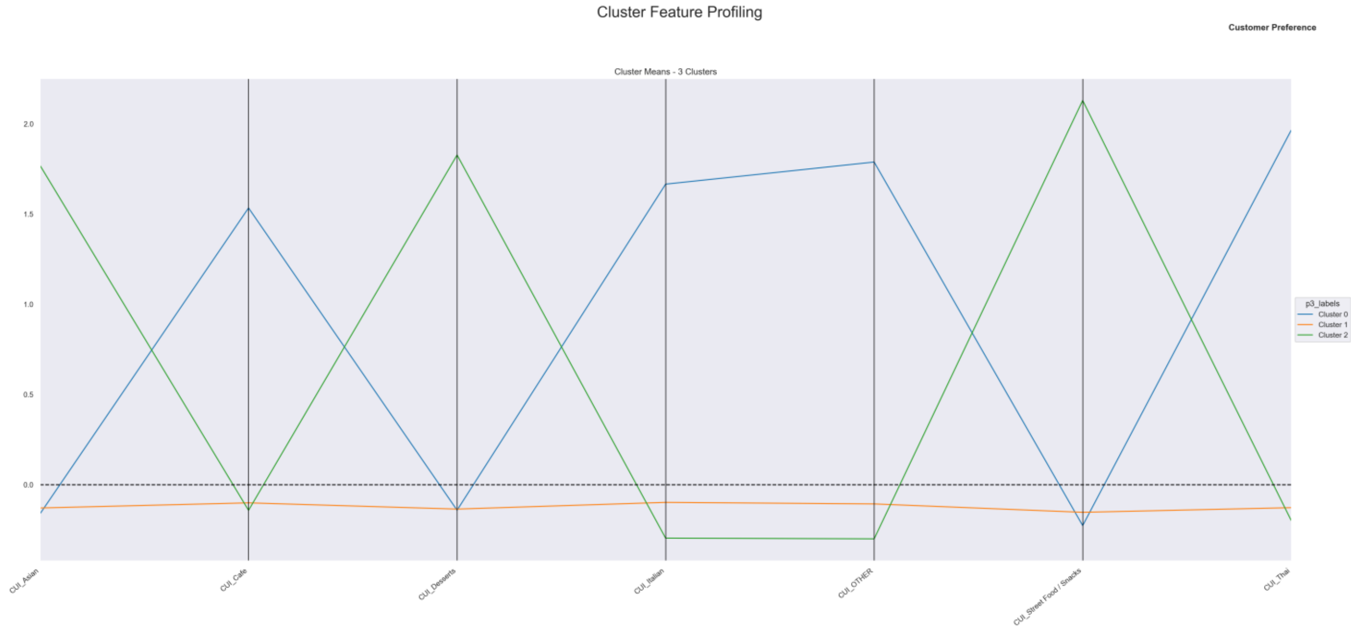

Figure 27 – Cluster feature profiling for perspective 3

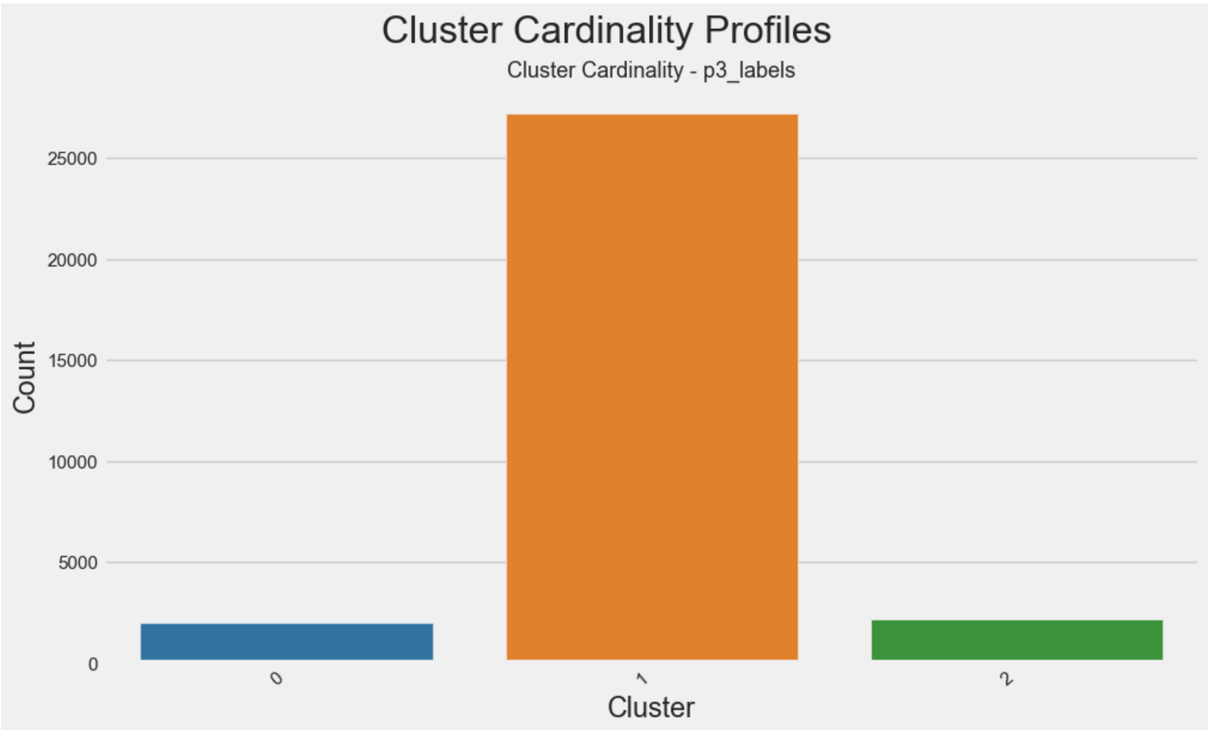

Figure 28 – Cluster cardinality for perspective 3

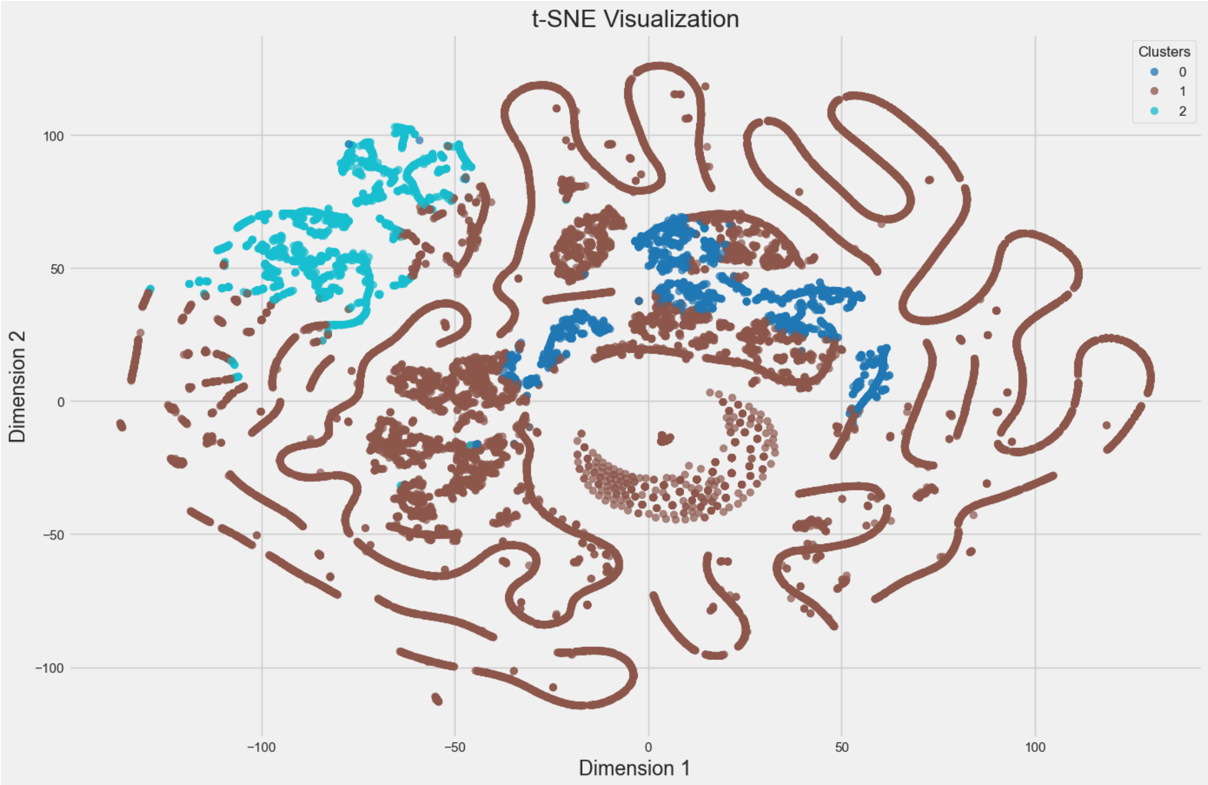

Figure 29 – t-SNE for perspective 3

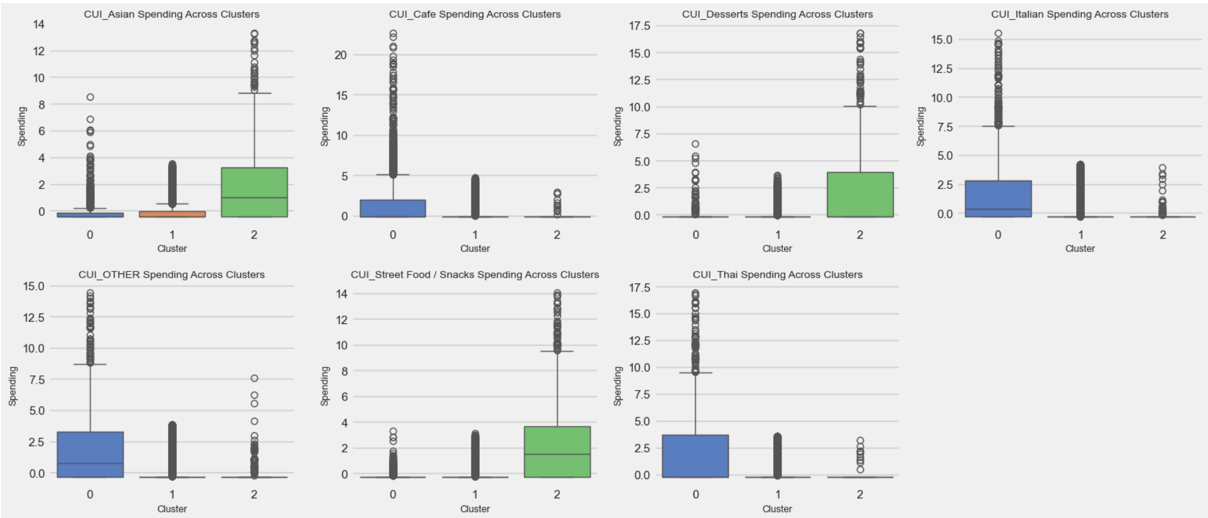

Figure 30 – Clustering features boxplots for perspective 3

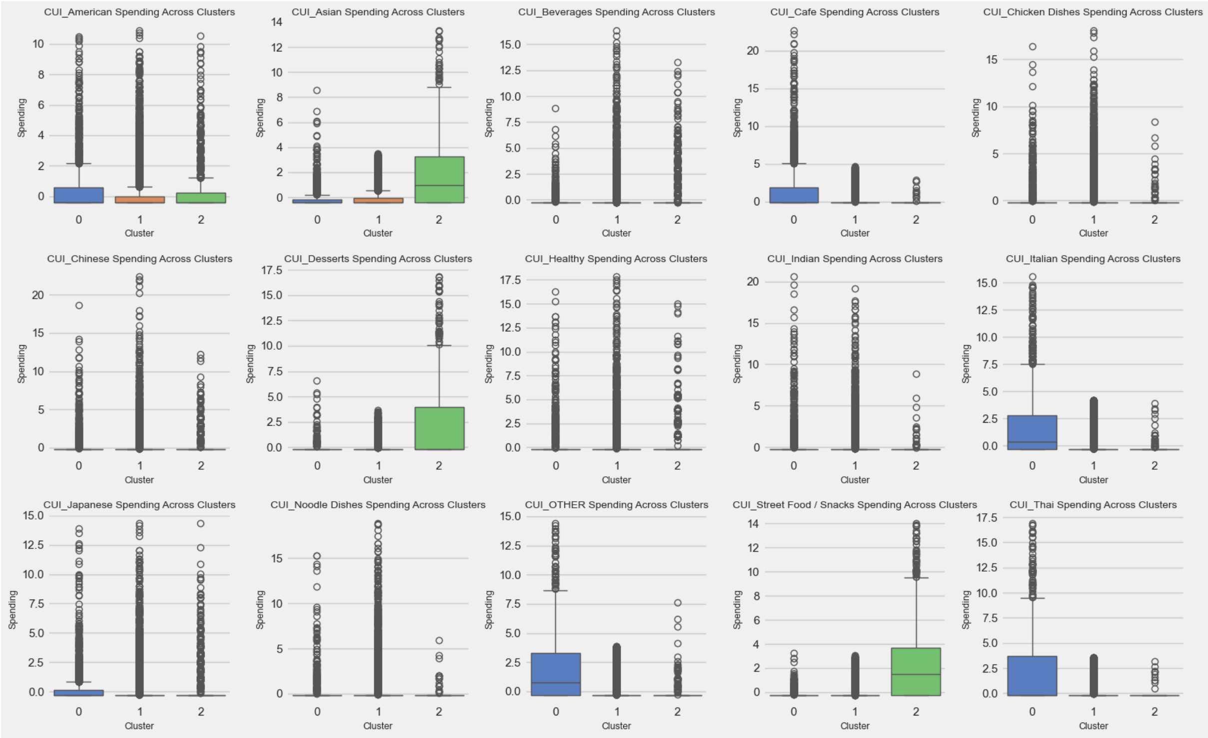

Figure 31 – Cuisine types boxplots for perspective 3

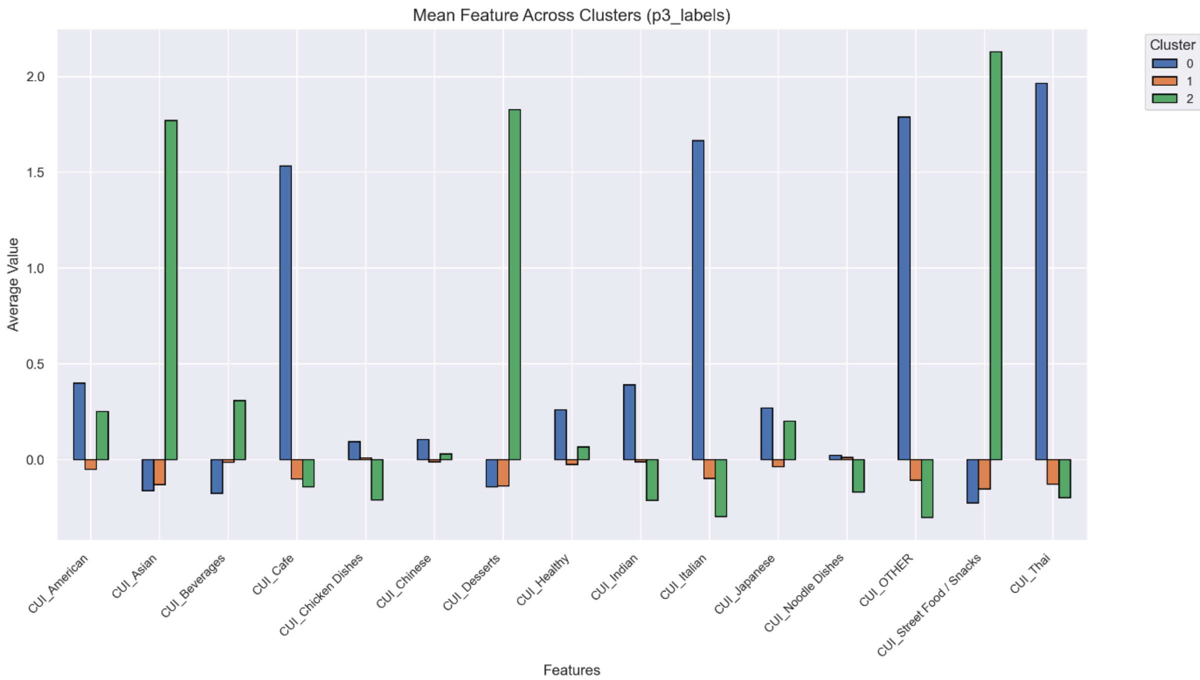

Figure 32 – Cuisine types bar plots for perspective 3

Figure 33 – Total spending boxplots for perspective 3

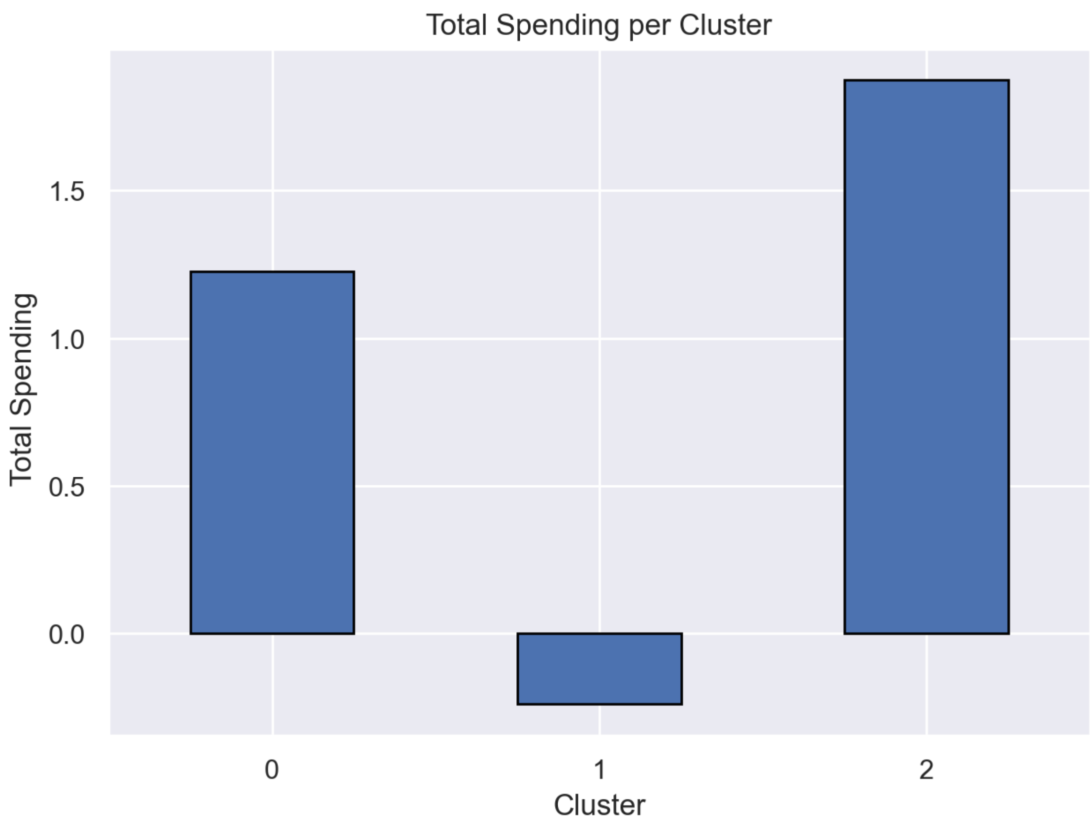

Figure 34 – Total spending bar plot for perspective 3

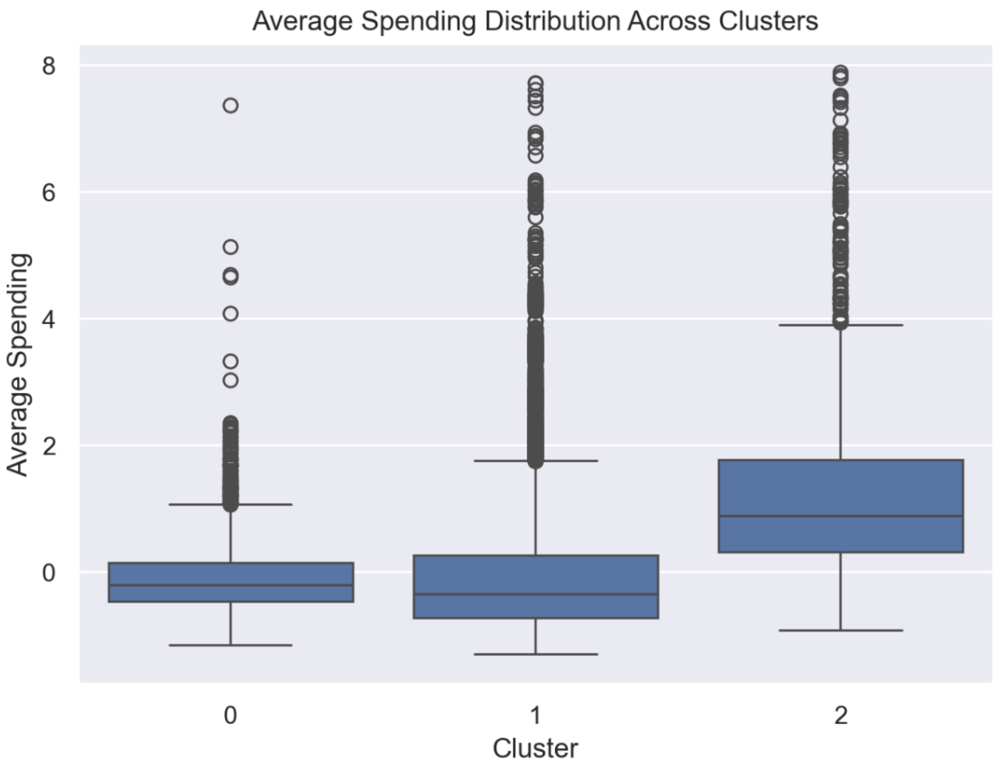

Figure 35 – Average spending boxplots for perspective 3

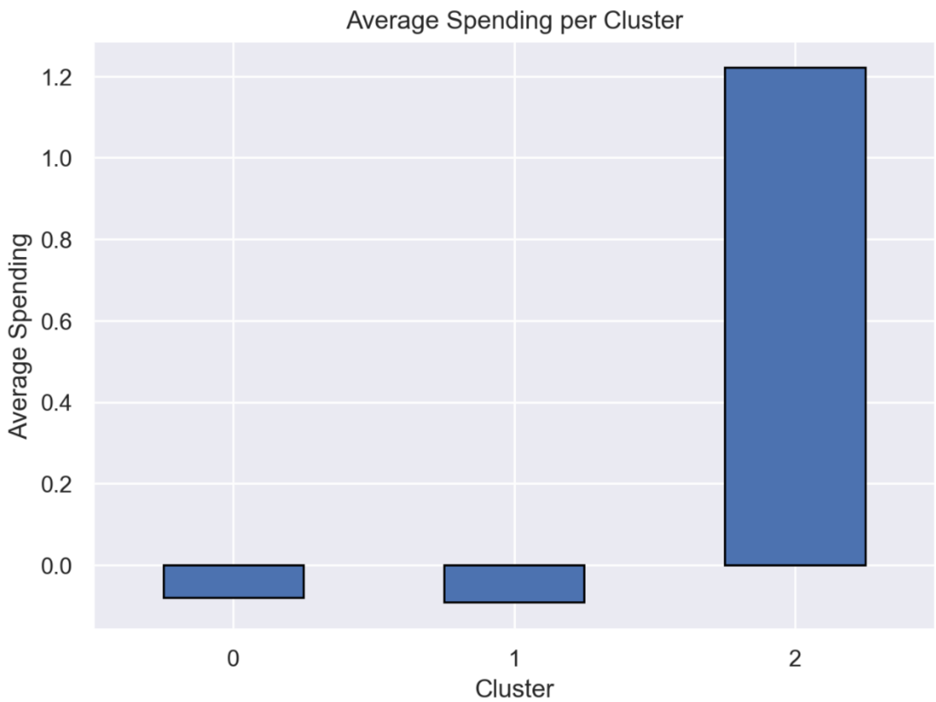

Figure 36 – Average spending bar plots for perspective 3

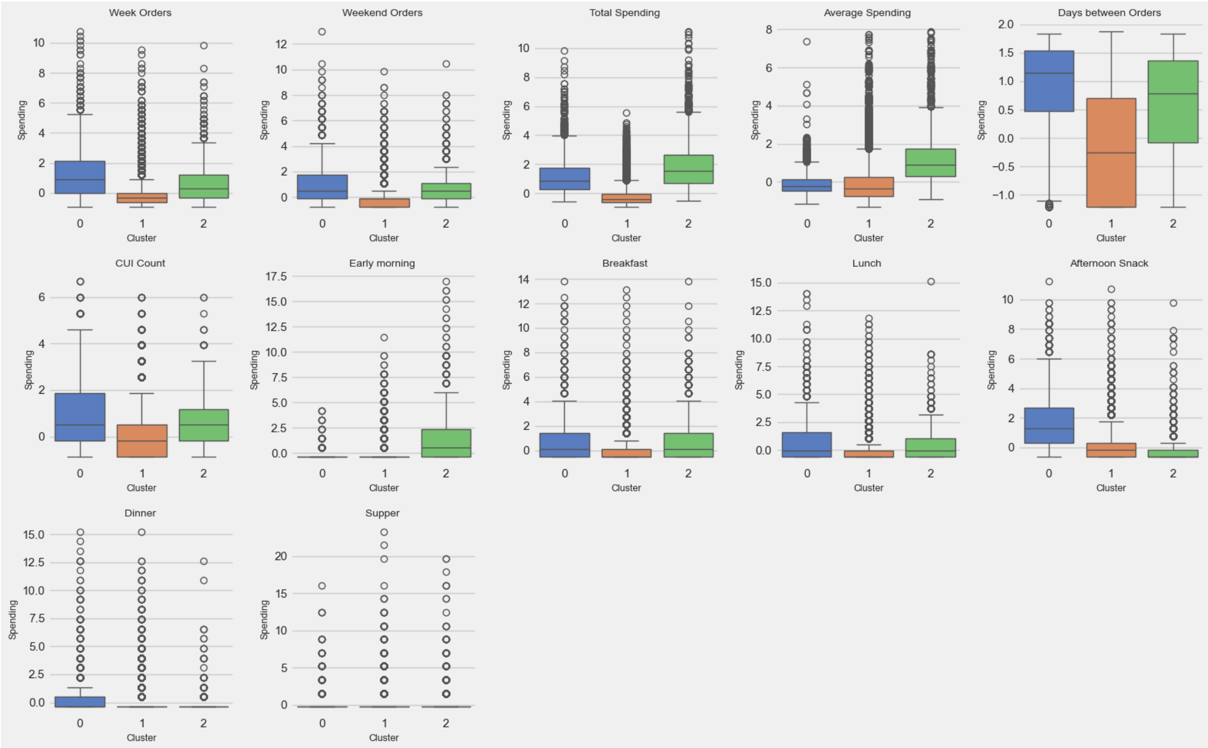

Figure 37 – Unused numeric features boxplots for perspective 3

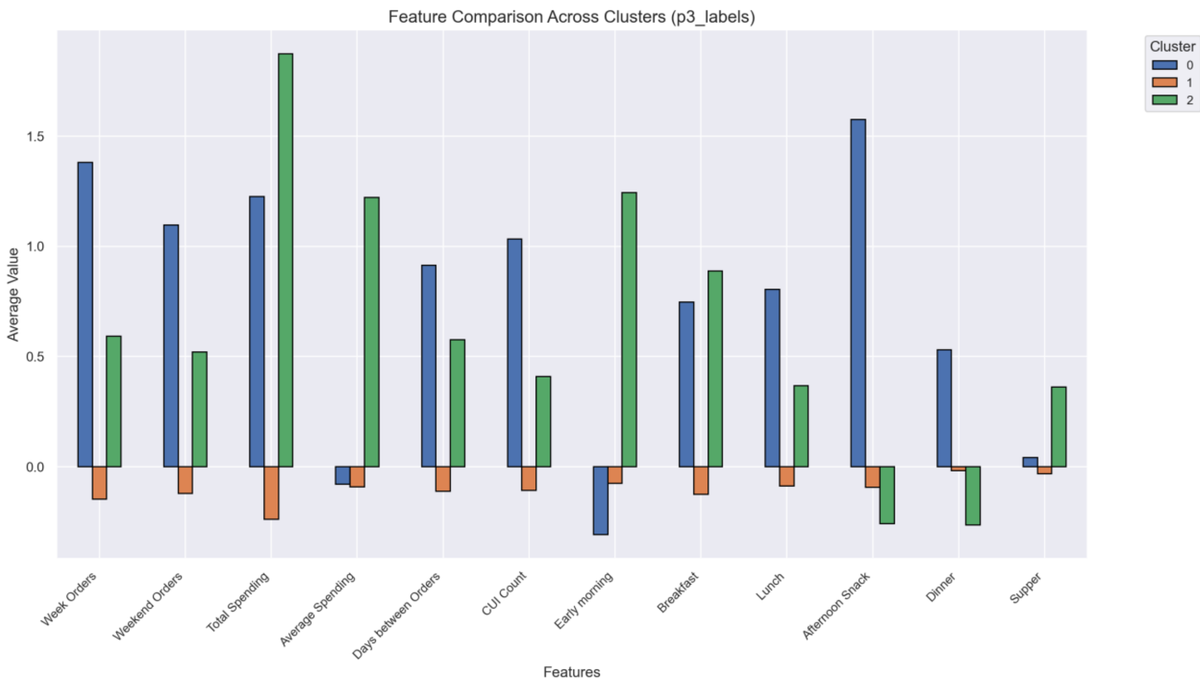

Figure 38 – Unused numeric features bar plots for perspective 3

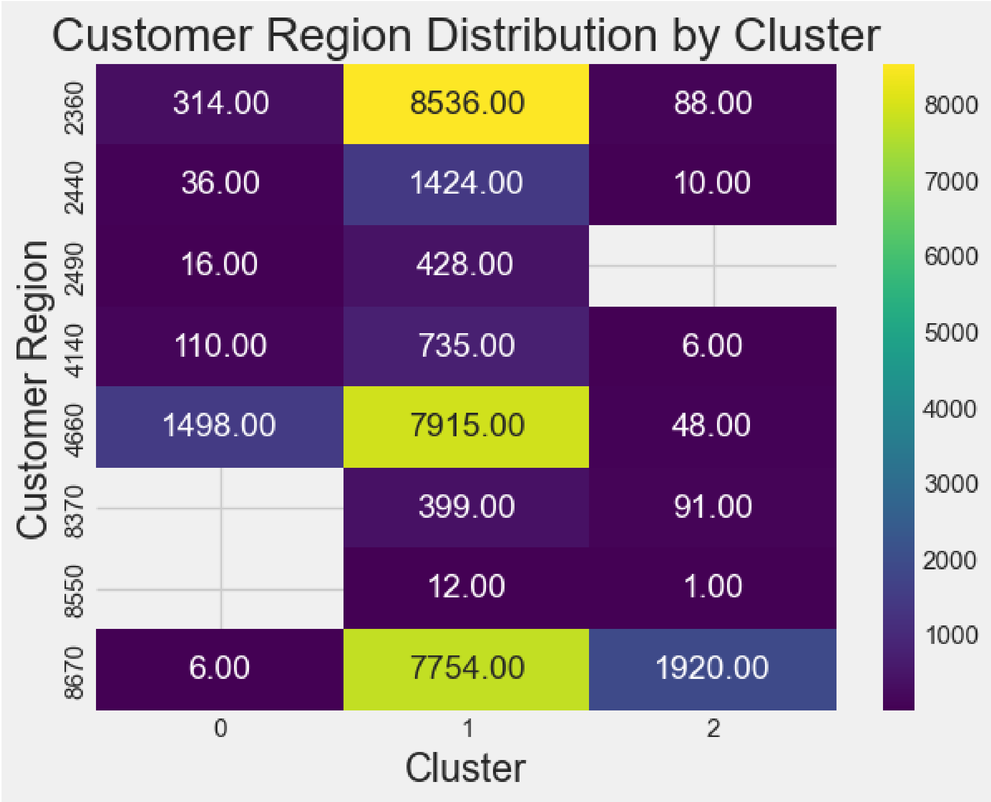

Figure 39 – Customer region distribution for perspective 3

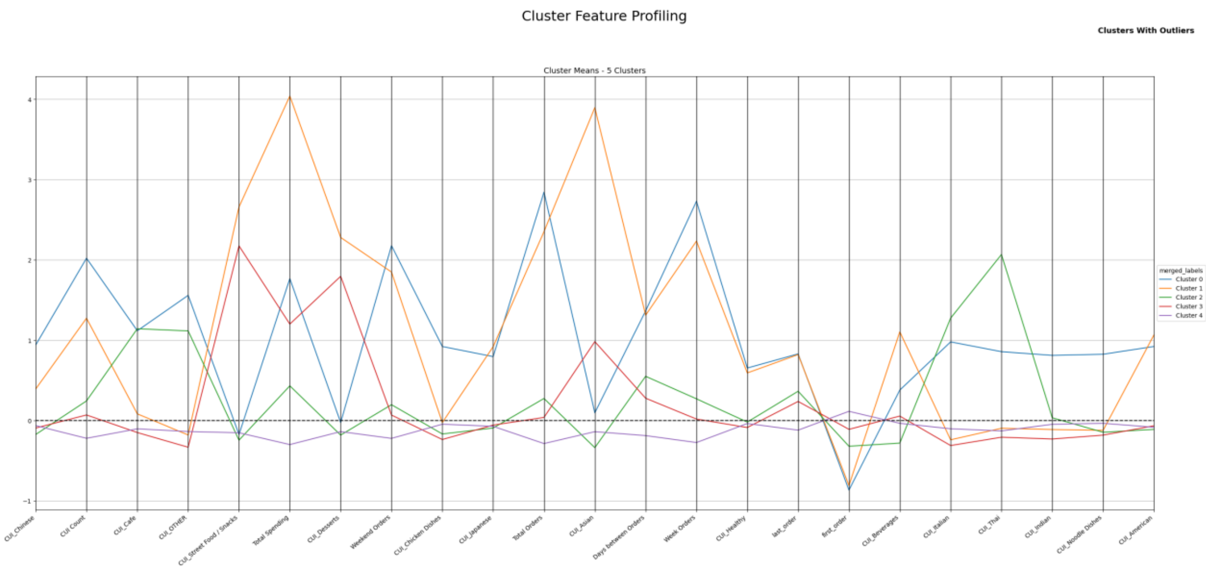

Figure 40 – Cluster feature profiling for final solution with outliers

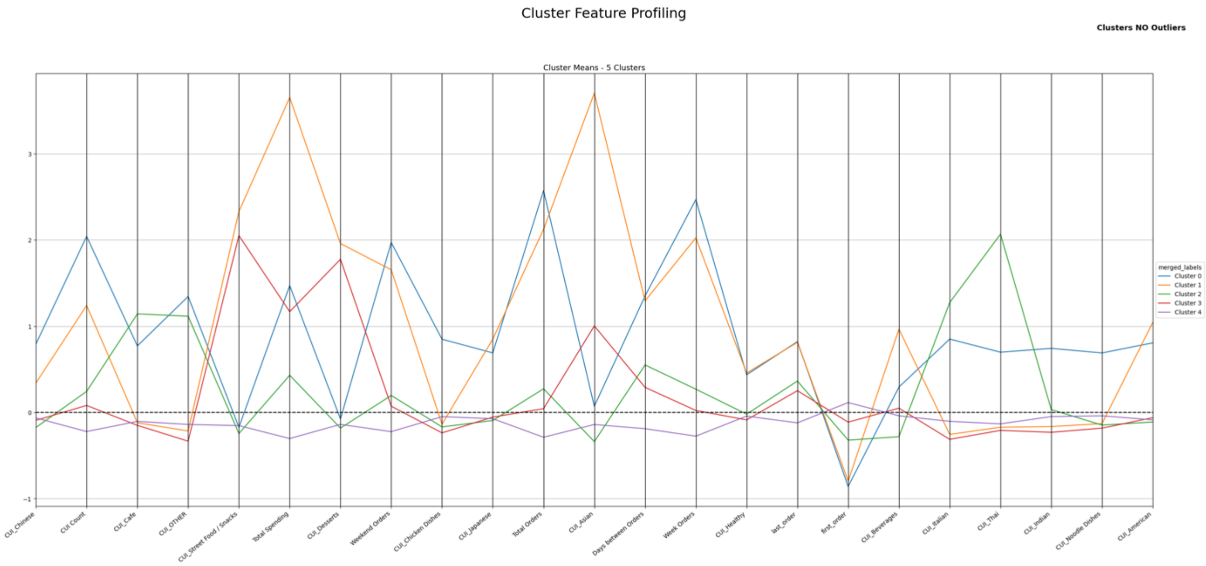

Figure 41 - Cluster feature profiling for final solution

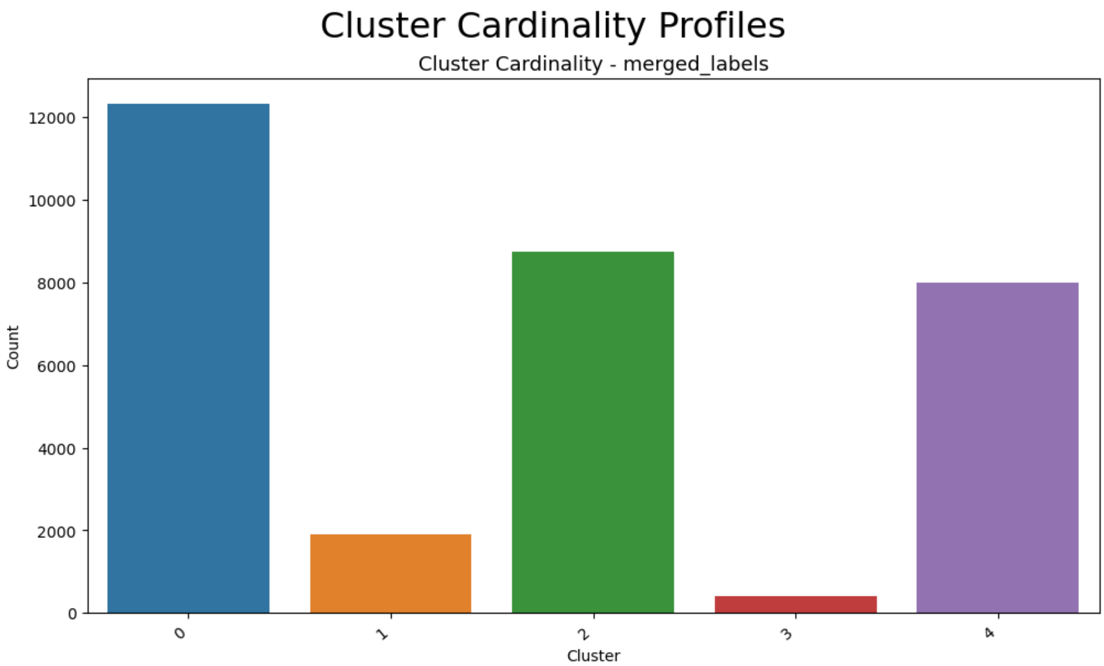

Figure 42- Final solution cluster cardinality

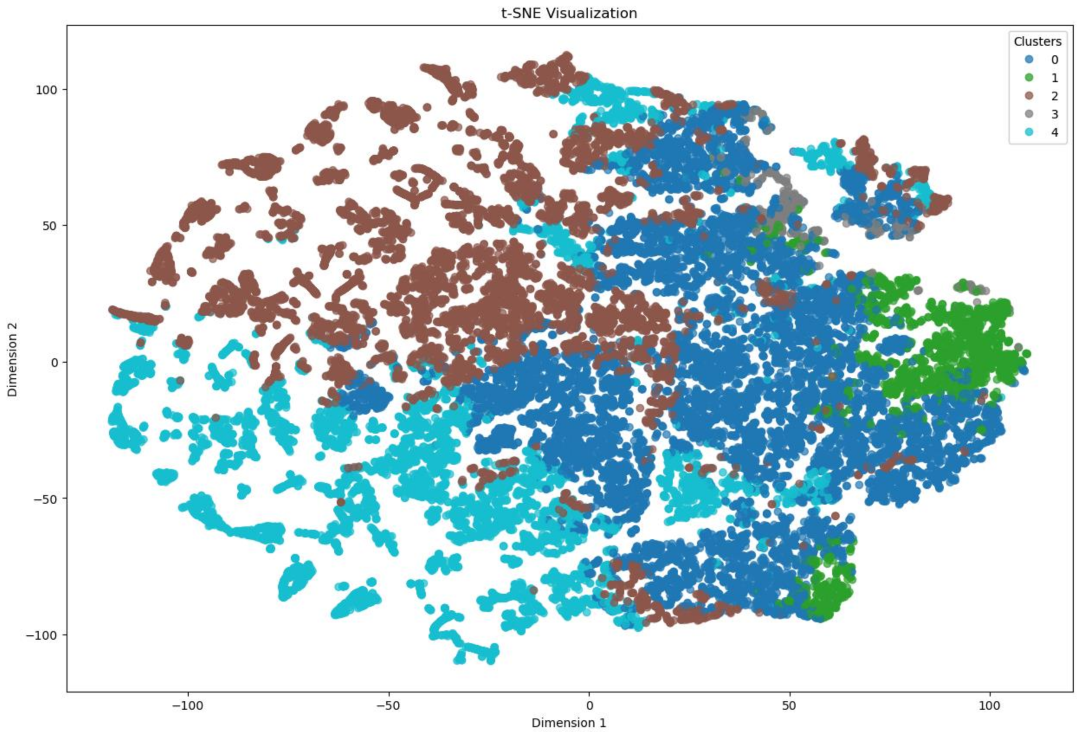

Figure 43 – t-SNE for final solution

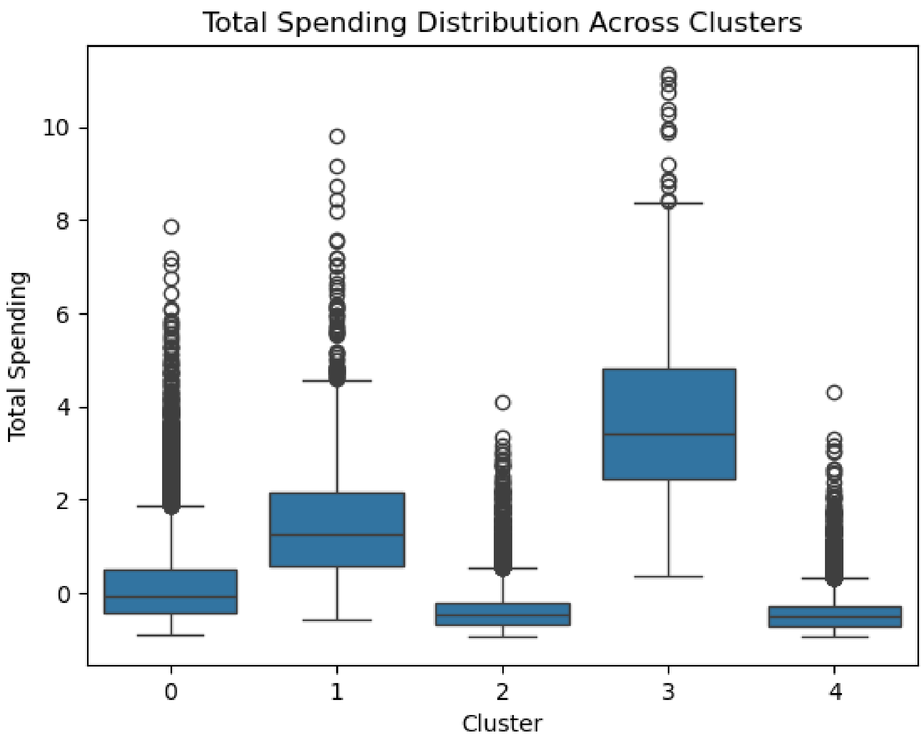

Figure 44 – Total spending boxplots for final

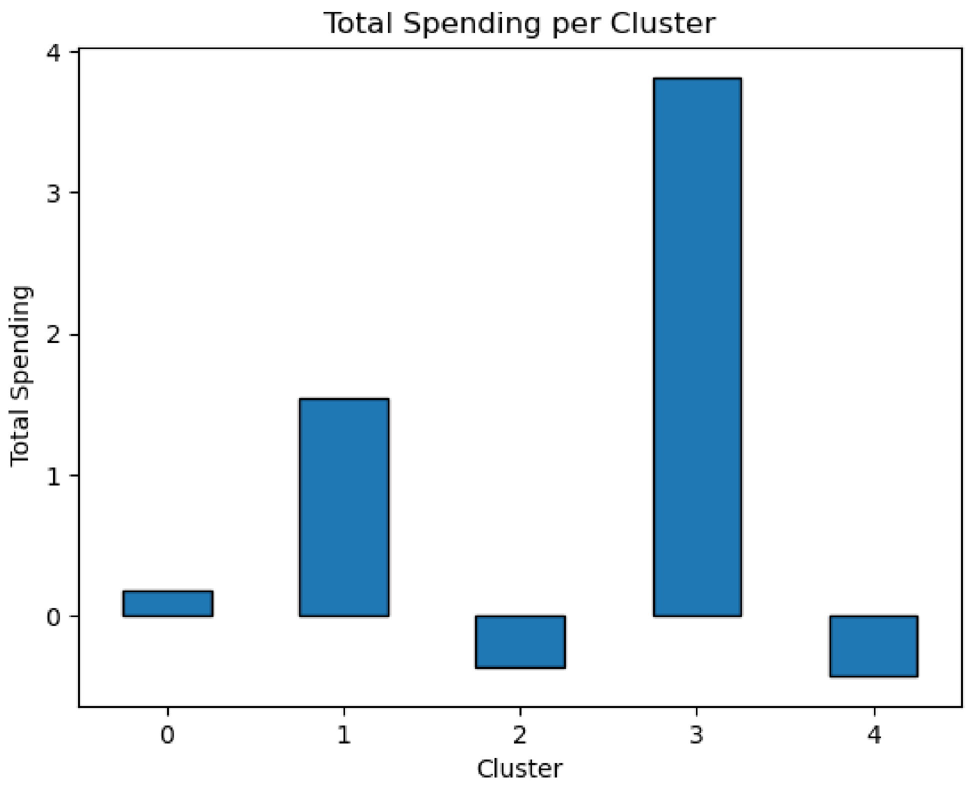

Figure 45 - Total spending bar plot for final solution

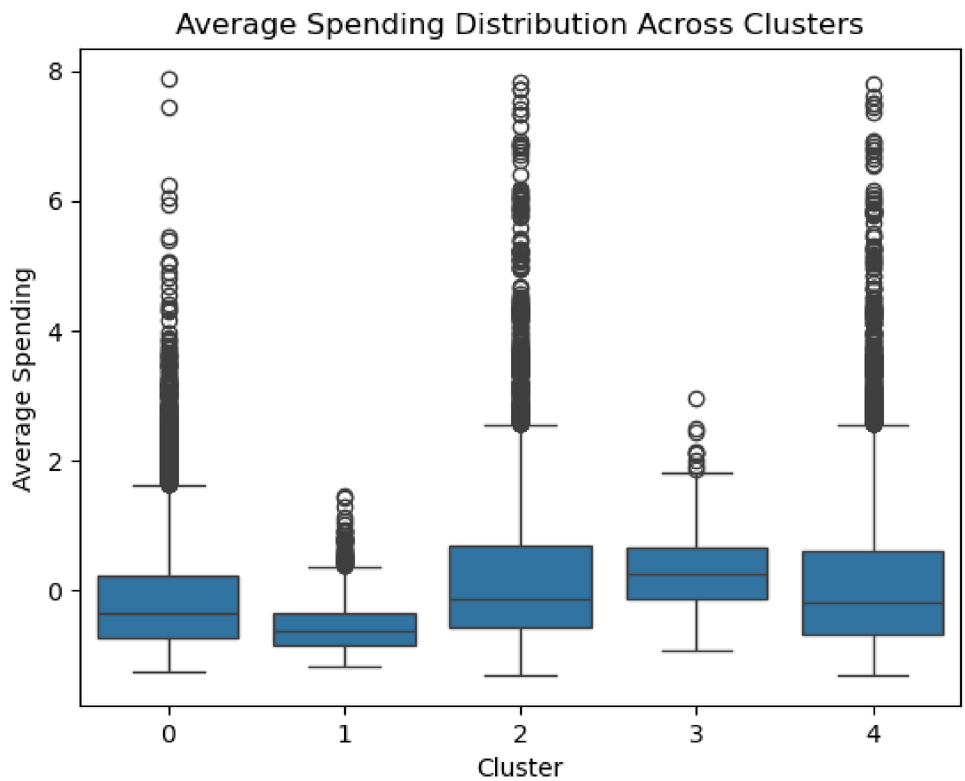

Figure 46 – Average spending boxplots for final solution

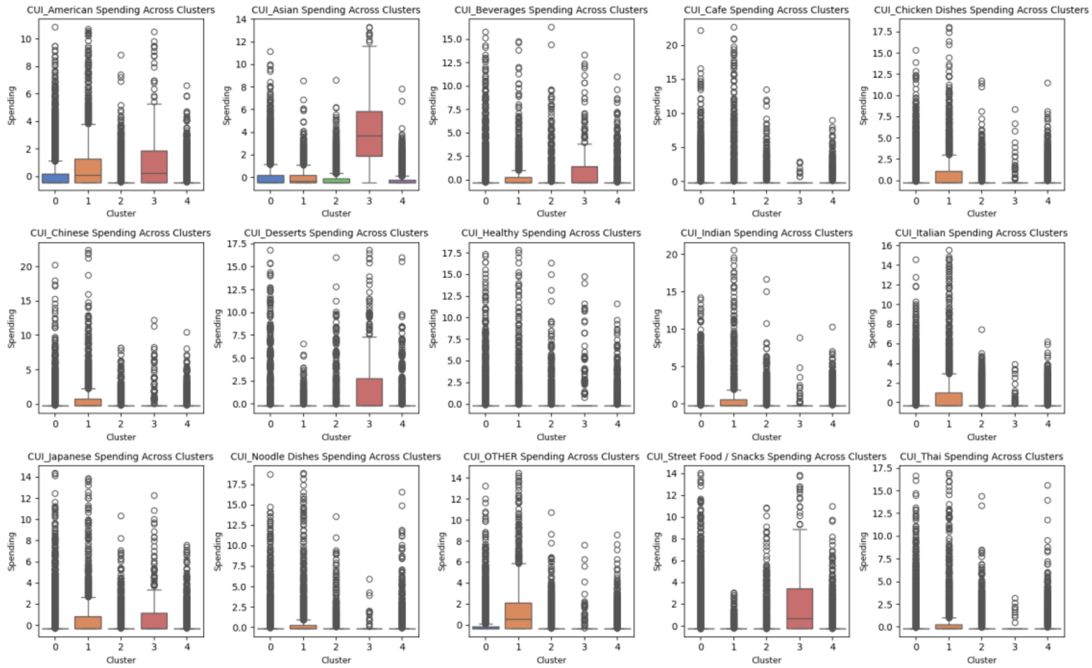

Figure 47 – Cuisine types boxplots for final solution

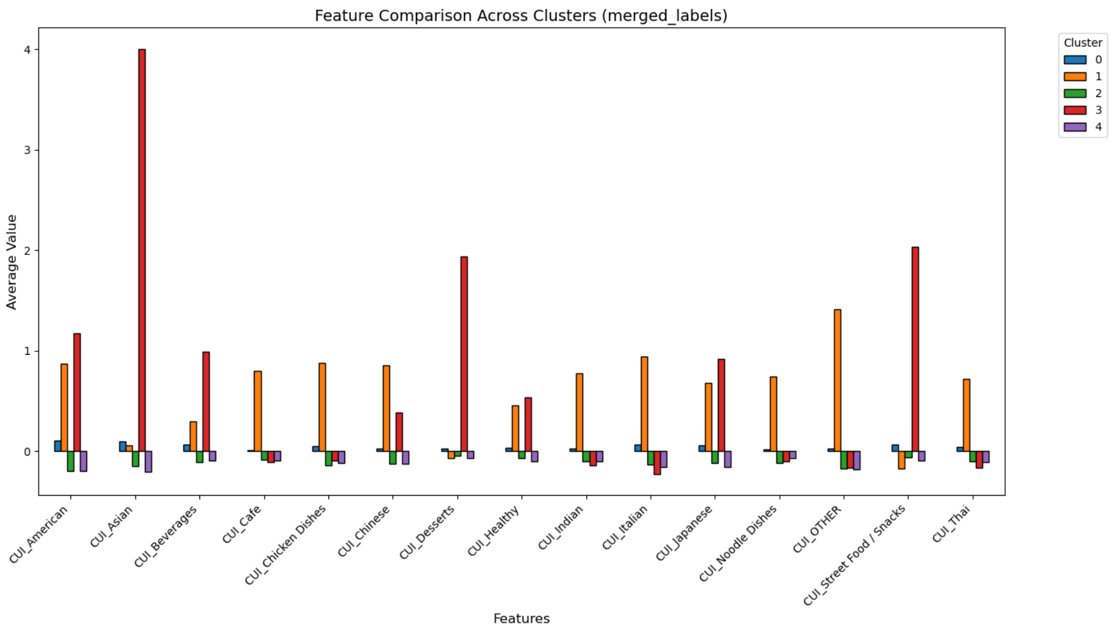

Figure 48 – Cuisine types bar plots for final solution

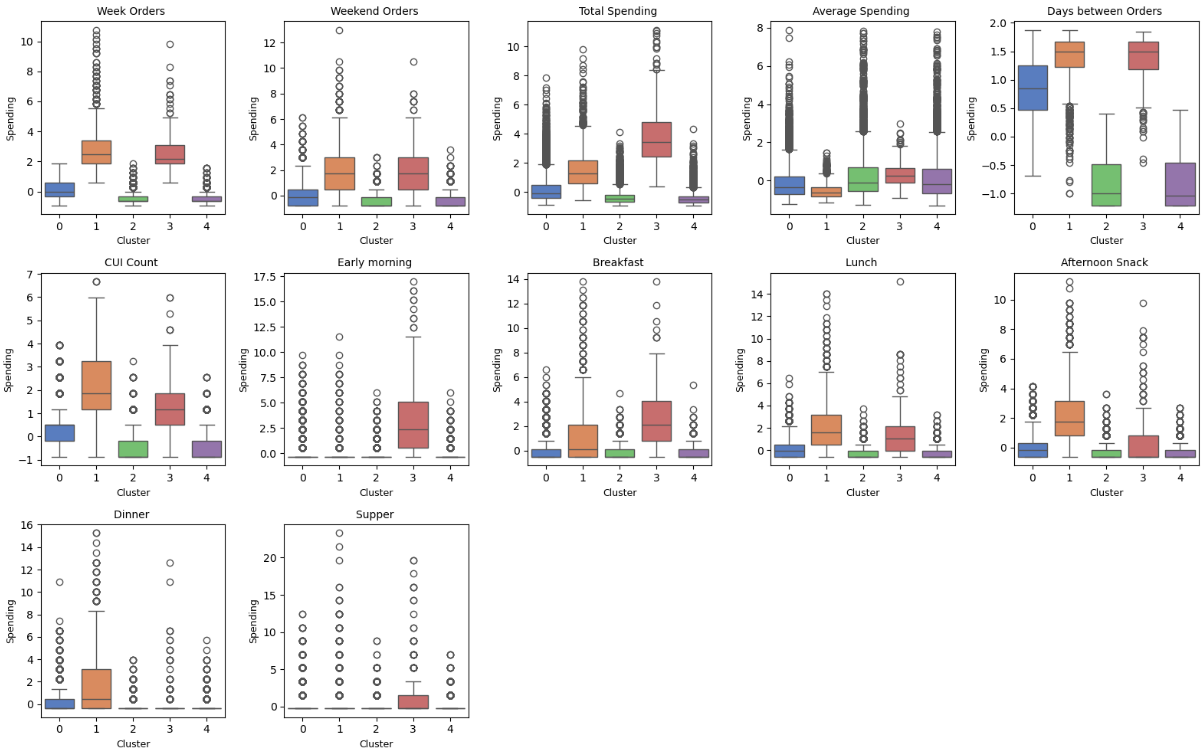

Figure 49 – Unused numeric features boxplots for final solution

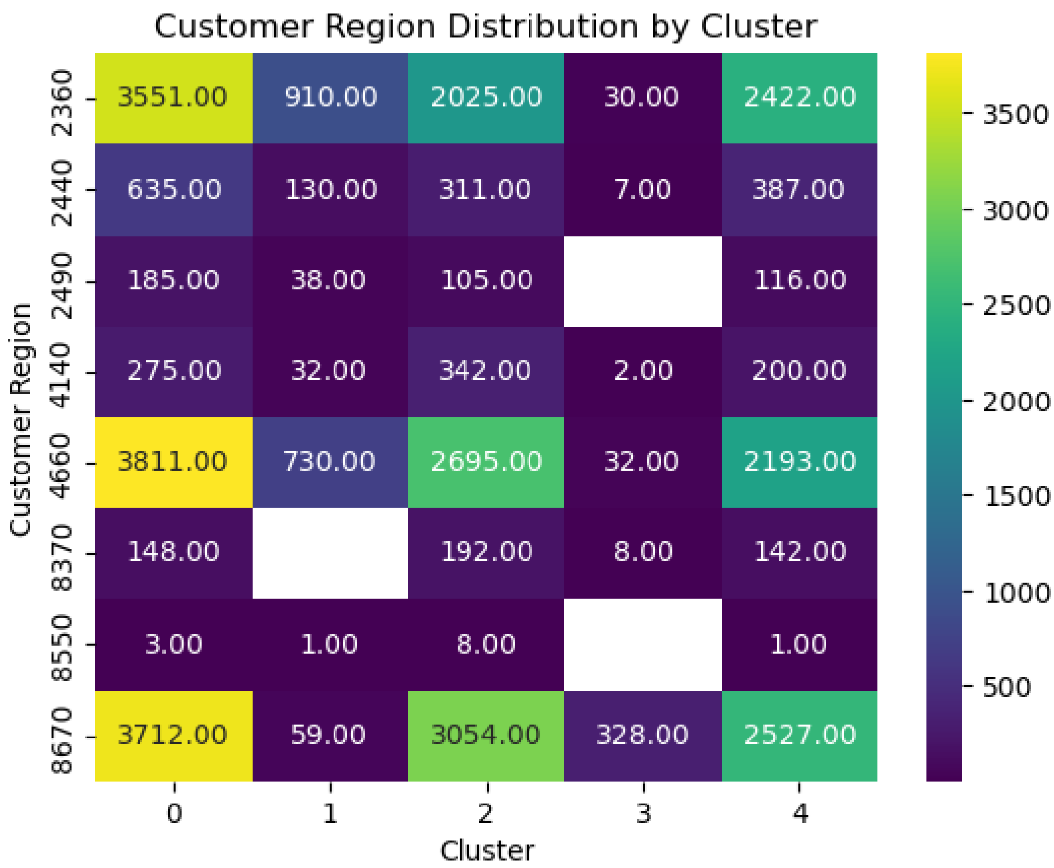

Figure 50 – Customer region distribution for final solution

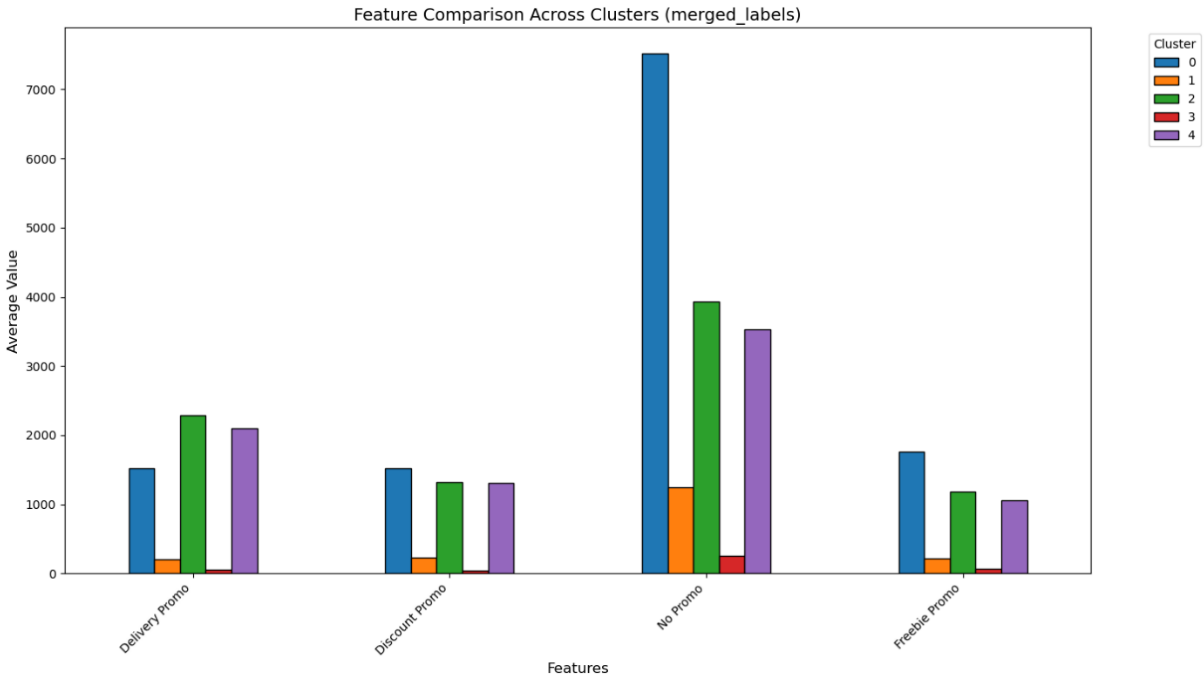

Figure 51 – Different promotion usage for final solution

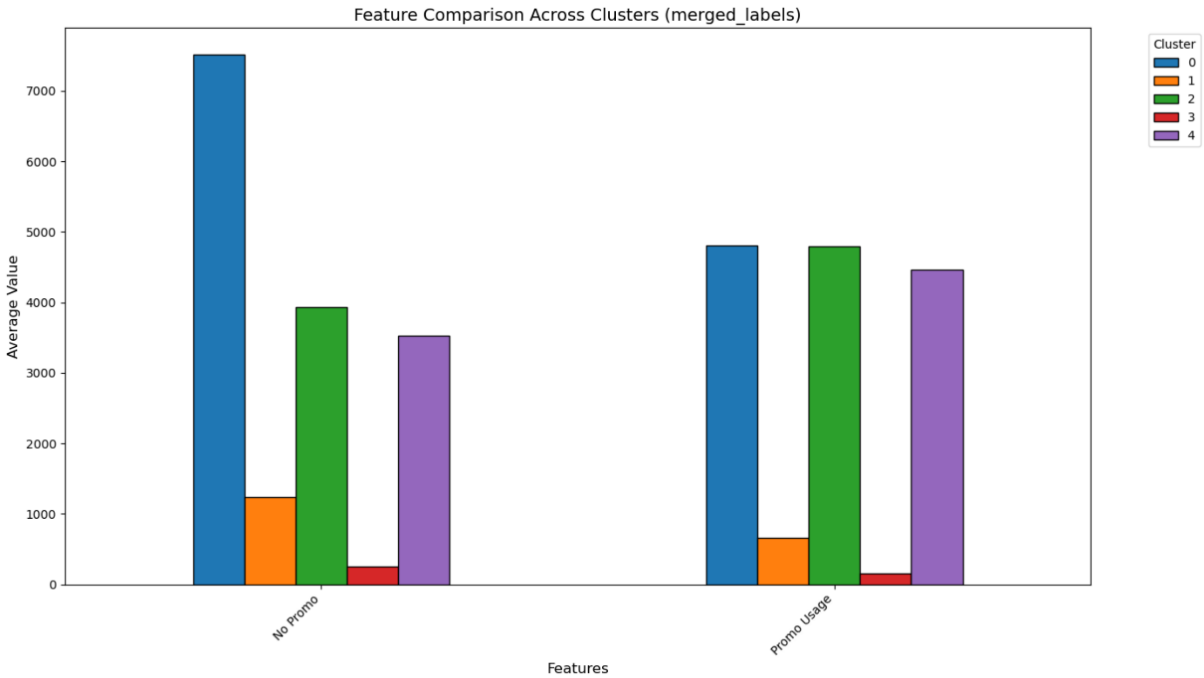

Figure 52 – Promotion usage for final solution

Figure 53 – Outlier Removal