NOVA IMS Information Management School

MDSAA

Master's Degree Program in

Data Science and Advanced Analytics

DATA MINING

Segmentation of Customers

Group 79

Diogo Melo, ID: 20240698

Hugo Trigueiro, ID: 20240577

NOVA Information Management School

Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

04 January, 2025

TABLE OF CONTENTS

ABSTRACT

This report explores customer segmentation for ABCDEats Inc., a fictional food delivery service. Using data collected over three months from three cities, we analyze customer purchasing behavior and demographic attributes to develop actionable insights. The recommended segmentation approaches include value-based, behavior-based, and preference-based. To achieve these goals, we employed advanced clustering techniques such as K-means, Hierarchical Clustering, and Self-Organizing Maps (SOM) across all segmentation perspectives. We evaluate the adequacy of our clustering solution by recurring to 3 different metrics: R2, Silhouette score, and the Calinski-Harabasz score. By integrating these perspectives, the study provides ABCDE with a data-driven framework to enhance customer satisfaction, retention, and revenue growth.

1. INTRODUCTION

Consumers are becoming increasingly selective about the businesses they support, making it critical for companies to gain a deep understanding of their customer base. For a food delivery company like ABCDEats Inc., identifying and addressing diverse customer priorities, whether they value convenience, affordability, or healthy options, is essential to delivering tailored services and gaining a competitive edge.

Customer segmentation is the cornerstone of this process, dividing a broad customer base into smaller groups with shared characteristics or behaviors. By doing so, businesses can customize their products, services, and marketing efforts to better meet the specific needs and preferences of each segment, ultimately enhancing satisfaction, loyalty, and revenue.

To achieve this, we utilized a data-driven approach, analyzing customer data collected over three months from three cities. We started by understanding and visualizing, cleaning, modeling and finally, clustering. This analysis incorporated a range of segmentation models and techniques. We applied a **value perspective** to group customers by their economic contributions to the business: their total spent, average spent per order, and last order. Next, we used a **behavior perspective**, leveraging clustering algorithms such as k-means to identify patterns in purchasing habits. Additionally, **preference perspective** was performed to classify customers eating preferences, the types of cuisines.

Finally, we tend to present some marketing strategies that the company can use based on the final segmentation we provided.

Link for GITHUB: DiogoMelo08/Data-Mining-Project

2. DATA EXPLORATION AND PREPROCESSING

In the second part of this delivery, we maintained most of the EDA, since we were satisfied with our data treatment in the first delivery.

However, we did change some things in this delivery such as:

In the variable "Customer Region" we determined the optimal number of clusters using the elbow method, 3, and by using KMeans model that used other features in our data set and the values were imputed based on the most common value within each cluster, providing what we consider to be a reasonable estimate for the missing values. Most of them were associated with region "8670", as we presumed in the first EDA regarding customers behaviour in Cuisine Preferences. see annexes (Fig 0).

2.1. Outliers

In the 1st delivery we identified that there were outliers and in this 2nd delivery we decided to remove them and made sure we didn't eliminate a lot of information with the overall percentage of the data removed being around 2% of the data.

In our initial features we found outliers in features is_chain (Fig 1), product_count (Fig. 2) and vendor_count (Fig. 3).

2.1.1. Outliers Treatment

We proceeded to remove the outliers "manually" since the data is too skewed and using z-score and IQR methods wouldn´t work since it would lead to the removal of too many data points, as we previously mentioned in the first delivery.

Outliers were removed based on the plot which led to filtering product_count at value 100 mark (Fig. 4), vendor_count at 35 (Fig. 5) and is_chain at 60 (Fig. 6).

In feature engineering we decided to remove outliers in features total_spent (Fig. 7) variable, total_orders (Fig. 8) and avg_spending_per_order (Fig. 9) based on the plot provided.

Outliers were removed based on the plot which led to filtering total_spent at value 620 mark (Fig. 10), total_orders at 70 (Fig. 11) and avg_spending_per_order at 90 (Fig. 12).

2.2. FEATURE ENGINEERING

In feature engineering, we decided to create new features, such as Cuisine Groups to be able to use Cuisines for clustering since there were too many features, and that would difficult the process of interpret them later due to too many variables.

We decided to group our cuisines based on general knowledge of culinary and what made the most sense to us, grouping the cuisines in the following manner.

Wester Cuisine: American and Italian Cuisines.

Asian Cuisine: Asian, Chinese, Indian, Noodle Dishes, Japanese and Thai Cuisines.

Complementary Cuisine: Beverages, Desserts, Cafe, and Healthy Cuisines.

Other Cuisines: Street Food/Snacks, Other and Chicken Dishes Cuisines.

We also made sure that the groups of cuisines were balanced in terms of overall value to lead to equilibrated cuisines and not have one type absurdly dominating over the others.

We also created two features to indicate the customers preferred cuisine and the customer preferred cuisine type.

Additionally, we **reorganized** our **HR groups** now being, HR_breakfast from 6 to 11, followed by, HR_Lunch_dinner from 12 to 14 and 19 to 22, HR_afternoon from 15 to 18 and HR_evening from 23 to 5.

Besides the points mentioned above, the EDA and feature engineering remained the same.

2.3. SCALING DATA

Why bother to scale the data?

Because without scaling, features with larger ranges can dominate the learning process, leading to biased models. Scaling ensures that each feature contributes proportionately to the model. StandardScaler was used to normalize numerical variables, where it transforms the distribution of each feature to have a mean of zero and a standard deviation of one.

We also considered using MinMax scaler, however, StandardScaler is relatively robust to the presence of outliers compared to MinMax scaling, as it relies on the mean and standard deviation rather than the range of the data.

OUR THREE PERSPETIVES:

To finely divide our customers into groups, so we could profile them into groups with clustering, we decided to create three different perspectives. We created these perspectives based on our own knowledge and based on the correlation between the features, seen in the heatmap correlation matrix (Fig. 13), so we wouldn´t use features with very high correlation with one another which would lead to the curse of dimensionality and unnecessary information.

We tried to keep the number of features as low as possible, without comprising the intent of the cluster, to maintain things simple and to have an easier time interpreting the clusters, as well as not to have too many data points.

PREFERENCES PERSPETIVE:

This perspective includes the types of Cuisines (groups) we created by joining several types of cuisines under one group, them being "Asian Cuisine", "Western Cuisine", "Complementary Cuisine", "Other Cuisines". With this cluster we intend to observe which group of customers prefers a specific type of cuisine and the relation to others. This might be valuable, for example, to know what kind of cuisines to recommend to customers.

BEHAVIOUR BASED PERSPETIVE:

This perspective includes the features that indicates the purchase behaviour in the realms of time and type of restaurant (chain or not), so we included the features "HR_morning", "HR_lunch", "HR_dinner", "HR_evening", "Weekdays" and "is_chain". Initially, we were using also “Weekend” feature, however, clusters obtained had similar behaviour with Weekdays, so we decided only to include weekdays. Due to poor performances on R2 variable “chain_order_ratio” we then replace that with “is_chain” original feature. This perspective perfectly aligns with preferred cuisines perspective since it´s also behaviour based. Even though we thought about having it as one feature, after experimenting, we concluded that it led to many features, which led to a very incomprehensible cluster, making the decision to separate them in a more correct way.

VALUE PERSPETIVE:

This perspective focus on monetary value being composed by three features "total_spent", "avg_spending_per_order" and "last_order". With these features we can understand the money realm of our clients as well as access their recency of purchases with last_order (high values indicate high recency since it´s the number of days since the start of the dataset until last time customer order), with this perspective being very RFM oriented in a manner.

3. CLUSTERING TECHNIQUES

To evaluate our clusters and perspectives we assess the main three models: (Kmeans, Hierarchical Clustering and Self Organizing Maps). We focus and explore those giving cluster profiles for each one. For each technique we evaluate each one of the perspectives, merging at the end to get the cluster analysis and profiling with association of categorical features. (For Hierarchical Clustering alone despite doing the same steps, we are not going to mention this deeply as the other 2 models, as HC is already incorporated in those other two techniques.

METRICS USED TO EVALUATE EACH PERSPETIVE:

1. **R^2 Score** measures the proportion of variance explained by the clustering. It provides an indication of how well the clustering captures the structure of the data. Range [-1,1].
2. **Silhouette Score** measures how similar an object is to its own cluster (cohesion) compared to other clusters (separation). Range [-1,1]. Values near zero indicate overlapping clusters.
3. **Calinski Harabasz Score** measures the ratio of the sum of between-cluster dispersion to the sum of within-cluster dispersion. It is also known as the Variance Ratio Criterion. The higher it is, the better.

3.1. KMEANS

The goal was to find certain groups based on some kind of similarity in the data with the number of groups represented by K.

In all three perspectives…

We started by checking both the optimal number of clusters to keep with an elbow plot and to determine the optimal number of clusters that best separates the data (Silhouette). Thus, the number of clusters was chosen based on those techniques. see annexes (Fig. 14) – Value perspective **example**. (didn´t insert all in here, but the selection was made just as here in other perspectives).

Then, we initialize method, calculated distance of each sample to respective centroid and associated cluster index for each sample, based on minimum distance.

We saw it was well distributed by each cluster and none of them was with few samples.

At the end, we recur to t-SNE visualization as it is particularly useful for visualizing clusters and understanding the structure of the data. (reduce to two dimensions). see annexes (Fig. 15) (Fig 16 and 17 for other perspectives).

	Value	Preference	Behaviour
Number of Features	3	4	6
Number of Clusters	4	5	4
R^2 Score	0.62	0.52	0.51
Silhouette Score	0.41	0.54	0.48
Calinski Harabasz Score	17013	8704	11135

Table 1 – KMeans Perspectives Details

3.1.2. K-means Merging perspectives. KMeans with Hierarchical Clustering

Since we used three perspectives and have a considered number of clusters for each one, we decided to agglomerate clusters using **Hierarchical Clustering**.

We recur to Dendrogram using Euclidean Distance and "Ward" Linkage as we noticed, it was always better (in this case). We considered five as a good number of clusters to use, balancing 3 factors (Dendrogram, R2 Score and adequate number of clusters to profile at the end). If, this number was higher maybe we were overcomplicating things, if too low we were losing some information.

See annexes fig 18.

Hierarchical clustering looks to find a middle ground through a top-down (divisive) or a bottom-up (agglomerative) approach to the analysis of group similarities. The output of this method is a dendrogram that groups clusters according to their degree of similarity or linkage.

3.3. Self-Organizing Maps (SOM)

Self-Organizing Maps are a useful technique to define clusters by employing unsupervised models on the trained network. In this model's process, the clusters are mapped back to the original dataset by assigning each record to the cluster associated with the nearest node, identified as the best unit.

In MiniSOM - A rule of thumb to set the size of the grid for a dimensionality reduction task is that it should contain 5*sqrt(N) neurons where N is the number of samples in the dataset to analyze.

Random Weights init initializes the weights of the SOM picking random samples from data.

Quantization Error returns the quantization error computed as the average distance between each input sample and its best matching unit.

Activation Response returns a matrix where the element i,j is the number of times that the neuron i,j have been winner.

Train Batch trains the SOM using all the vectors in data sequentially.

Upon completing training, the key statistical features of the input space are displayed in lower-dimensional visualizations. These visualizations, in theory, preserve the essential patterns (i.e., the topological differences between neurons) of the high-dimensional data.

The first critical decision was to select map size and lattice for nodes. We decided to use a hexagonal lattice for its ability to represent spatial relationships more evenly. A network featuring a 20 by 30 node arrangement emerged as the most effective and reduced time consumption solution.

In all **three perspectives** we applied **Kmeans or Hierarchical Clustering on top of the MxN units**. For that goal, we plot Elbow method and Silhouette (Kmeans) and Dendogram for Hierarchical Clustering.

	Value	Preference	Behaviour
Number of Features	3	4	6
Number of Clusters	4	5	4
R^2 Score	0.52	0.42	0.47
Quantization error	0.18	0.17	0.30
Silhouette Score	0.32	0.53	0.56
Calinski Harabasz Score	11614	5753	9305

Table 2 – MiniSOM Perspectives Details

For each one of the perspectives, we also did a TSNE to visualization and also did a box plot for each feature of that perspective to visualize how points were distributed by each cluster.

From Figures 19 to 24 - SOM Value in appendix, we have additional visualizations from where we can see patterns regarding value perspective, since TSNE, to boxplots for each label, the mean of each feature for each label, hexagons of SOM HC and even association with Customer Region.

From Figures 25 to 30 - SOM Preferences in appendix, we have additional visualizations from where we can see patterns regarding preferences perspective, since TSNE, to boxplots for each label, the mean of each feature for each label, hexagons of SOM HC and even association with Customer Region.

From Figures 31 to 36 - SOM Behavior in appendix, we have additional visualizations from where we can see patterns regarding Behavior perspective, since TSNE, to boxplots for each label, the mean of each feature for each label, hexagons of SOM HC and even association with Customer Region.

5. CLUSTER ANALYSIS AND PROFILING.

Despite doing Cluster Analysis and Profiling for both techniques (Kmeans with HC and SOM with HC), we decided to use the best one in terms of R2 and Silhouette Score on merged clusters.

Option 1: Kmeans + HC had scores (R^2: 0.29, Silhouette: 0.18)

Option 2: SOM + HC had scores (R^2: 026, Silhouette: 0.55)

We **chose option 2 (SOM + HC)** due to R^2 scores being very similar, however **Silhouette** had a **huge** difference indicating that on **KMeans** clusters could be **overlapping** since the value was near 0. On **SOM**, with **0.55** (closer to 1) seemed the **best** option.

We already mentioned the SOM technique, the procedure for merged perspectives, so we will just show some resume of final clusters to understand it better and make decisions: (UNSCALED DATA) for better understanding.

labels	total_spent	avg_spending_per_order	last_order	Weekdays	is_chain	HR_breakfast	HR_lunch_dinner	HR_afternoon	HR_evening	Western_Cuisine	Asian_Cuisine	Complementary_Cuisine	Other_Cuisines
0	178.900261	6.273359	85.783290	21.305483	21.966057	10.749347	8.953003	9.261097	0.451697	33.644726	72.163290	33.654021	39.438225
1	31.333889	10.083077	62.799457	2.660040	2.435609	0.976311	1.131862	1.221913	0.415698	7.592330	14.000896	3.647702	6.092962
2	212.369690	28.568882	77.193798	6.906977	4.015504	3.503876	2.325581	1.813953	1.860465	8.550310	41.552016	7.042403	155.224961
3	146.813877	16.936297	78.744493	8.035242	7.070485	4.806167	2.004405	1.775330	2.859031	10.959956	26.777048	92.501454	16.575419
4	170.913774	15.618313	82.532075	8.943396	6.211321	4.197484	2.480503	2.564780	3.328302	15.070918	116.521107	14.014403	25.307346

Figure 37 – Mean of each Merged Clusters (using Kmeans and HC)

Cluster analysis for visualization can be seen at appendix (FIG 38)

Boxplots of metric features after merged clusters at appendix (Fig. 39)

Zero Cluster overall: Spends a lot, with a low average for orders and their last time was recent. They spend a lot during weekdays, order a lot from chains restaurants and the only time slot where they spend less is in the evenings with the 2nd lowest value. Overall, in preferences, they have a solid behavior in all cuisines, however, comparing with other clusters, this group has the highest spent in Western Cuisine.

We can possibly tell that these are individuals who the company should value a lot due to its cluster overall performance. 91% of these customers are associated 91% with 2 regions (2360 and 4660), they preferred to pay with card (77%). 70% of them didn´t use any promo in last order and have most young adults (biased feature, as it has a lot of young adults).

We can presume that these are workers, who are single (marital status), don´t have kids or just like to eat alone, (to justify low average spending per order) comparing with the other high values in other features such as total spent. These people don´t seem to buy at all at night, so we can presume night life it´s not very frequently or people enjoy earlier meals.

Marketing strategy – Offer discount for two meals (example: 2 Asian dishes) to psychology tell "ok, if I get one more or say to my coworker/other to join me, we will both enjoy discount. Other times, mix marketings, we can also attribute points to each order, and when it gets to 10, a free meal his offer, since these people usually order a lot of times, it´s a good practice to feel them rewarded. Also, they spent a lot in chain restaurants, so promos or new chain restaurants should be mentioned to this group.

1st Cluster overall: Our lowest value customers are in this group. They have a very low total and average spent, are the group with lowest recency (last come they order since start of dataset was the lowest). They don´t have a preferred time slot and they preferred type of cuisine is Asian and Western also. It seems that this group has the highest proportions in used promos, meaning that they could be waiting for more promos to order.

The regions are equally distributed between (2360, 4660 and 8670), and they have the largest proportion in paying with card and cash. Young adults are the most dominant in this group. However, teens and mid adults have the highest proportion considering other groups.

Marketing strategy: We can also use the 10 points (1 per order) to gain the attention of these customers. This would be a way of awake inactive customers and trying to make them order more with a goal, since they look mostly for promotions. This would increase the overall performance of this group.

2nd Cluster overall: This group has the highest total and average spent. Their average recency is also high (which is good). This group doesn´t consume the usual at most, as they preferred type of cuisine is Other (Street Food & Snacks, Chicken and Other cuisines). They order a few times from chain restaurants, and they consume the most at breakfast and lunch/dinner.

The proportion of mid/old adults and senior is bigger than other groups. 83% of customers in this group are from region 8670. These customers have the highest average spending, indicating they are willing to spend a big amount of money to get what they want. We can associate this group with people who like other types of cuisines rather than the most usual ones, this indicates also more adventurous people. They also consume a lot of chicken and Asian cuisine.

Marketing strategy: Advertise more complementary cuisines (coffee shops like Starbucks, or good Donuts/Desserts Shops) to this group as we believe it can help to increase sales, due to its different aspect of creativeness and more mature things.

3rd Cluster overall: This group is a middle one in terms of value perspective. Their total and average spent so as the recency shows that they are in middle terms compared with other clusters. They usually order on weekdays and from a chain restaurant and they spend the most at breakfast and late evening. They consume a little bit of each cuisine, however, when it comes to Complementary Cuisine (Beverages, Desserts, Coffee and Healthy), they are the top spenders.

The proportion of old adults/ seniors it practically null, and they are more from region 4660 and 8670.

Marketing strategy: We can do something like the previous cluster but advertise also identical types of cuisine such as Other (Street Food & Snacks and others). For instance, advertise that a restaurant/store is offering breakfasts buffets or a box with different cakes/sweets at a good price for time limit.

4th Cluster overall: This group also has a good total and average spent on the business and their recency is high. They are the ones that consume the most Asian Type of Restaurants. Regarding time slots, evening is also the most they consume compared with other groups.

Most of this group is also from region 8670.

Marketing strategy: Due to the lower spent on lunch/dinner time slots, we could try to make promos for deliveries in these time slots to increase.

Some of these visuals come from the figure above (mean of feature per label), but sometimes we mention other figures such as association with categorical features. Such as Figures from 40.

6. SUGESTIONS:

We would also like to make some suggestions or procedures that the company could do to maximize their profit and make clients the most rewarded/happy:

1. See **inactive customers** (more than 30 days without ordering, for instance), what they order for the last 2 times or what it is their favourite cuisine → email marketing and get feedback.
2. Do **inquiries by email**, asking customers what restaurant or type of food they would like to see available on app. Company, then can try to reach those type of restaurants and do partnerships after testing profit scenarios.
3. If a restaurant establishes a relationship with ABCDEats, company should group that restaurant into a specific type of cuisines and **recommend that new restaurant** for people who **preferred that type of cuisine**.
4. Do **special promotion** for **birthday** in whatever they like to order (**like 30% discount**).
5. Observe **usual time slot of the day** or day of the **week** and **remember** customers by **SMS** that their **food** is **waiting** for a **click** to be delivered.
6. Offer **special promotions** during **holidays** or **seasonal** events. For example, a **summer** discount on cold **beverages** or a **winter** promotion on **hot meals**.
7. Allow customers to leave **reviews and ratings for restaurants and** dishes. Highlight top-rated items and restaurants to build **trust** and **attract more** orders. (**Example: The Fork in Portugal or UberEATS).**
8. **Actively engag**e with customers on **social media platforms** by sharing updates, promotions, and responding to customer inquiries. This helps build a community and increases brand **loyalty** – we can see the age category and adjusted type of publication and in what social media platform. (**Facebook for more older people**).
9. Implement a loyalty program where customers earn points for each order, which can be redeemed for discounts or free items. This encourages repeat business and rewards loyal customers. (**ON APP, for instance TOMATINO or McDonald´s**)

7. Conclusion

In this study, we understand that less perspectives or focus more on one set of features regarding company goals of segmentation could improve the result. Since, each one of the perspectives performed well alone, however when we merge clusters, it will increase noise, and if don´t have more clusters, the final solution could lead to lose of information.

We believe our perspectives were well defined and that the company could choose one of those based on their ultimate goals. We also suggested some improvements to increase profit, establish loyalty and make clients feel rewarded.

8. References

Farina Practical Notebooks NOVA IMS 2024/2025.

9. Appendix

customer_region	CUI_American	CUI_Asian	CUI_Beverages	CUI_Cafe	CUI_Chicken Dishes	CUI_Chinese	CUI_Desserts	CUI_Healthy	CUI_Indian	CUI_Italian	CUI_Japanese	CUI_Noodle Dishes	CUI_OTHER	CUI_Street Food / Snacks	CUI_Thai
-	5.87	21.62	4.84	0.00	0.00	0.85	2.23	0.88	0.00	0.00	3.54	0.00	0.03	10.42	0.00
8370	4.15	19.78	3.31	0.00	0.00	1.00	2.56	0.79	0.00	0.00	1.65	0.00	0.00	10.10	0.00
8550	12.81	22.58	8.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.40	0.00	0.00	10.93	0.00
8670	5.67	22.98	4.75	0.00	0.00	1.17	2.22	1.13	0.00	0.00	3.61	0.00	0.10	10.81	0.00

Fig 0 – Mean spent by similar regions to Unknown by each Cuisine

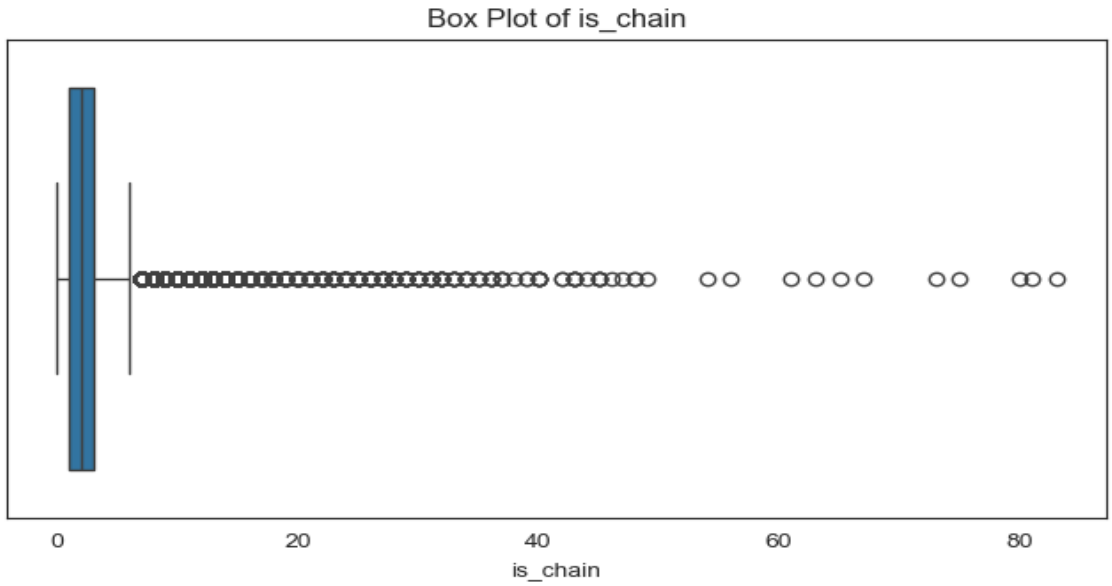

Figure 1 Box Plot of is_chain

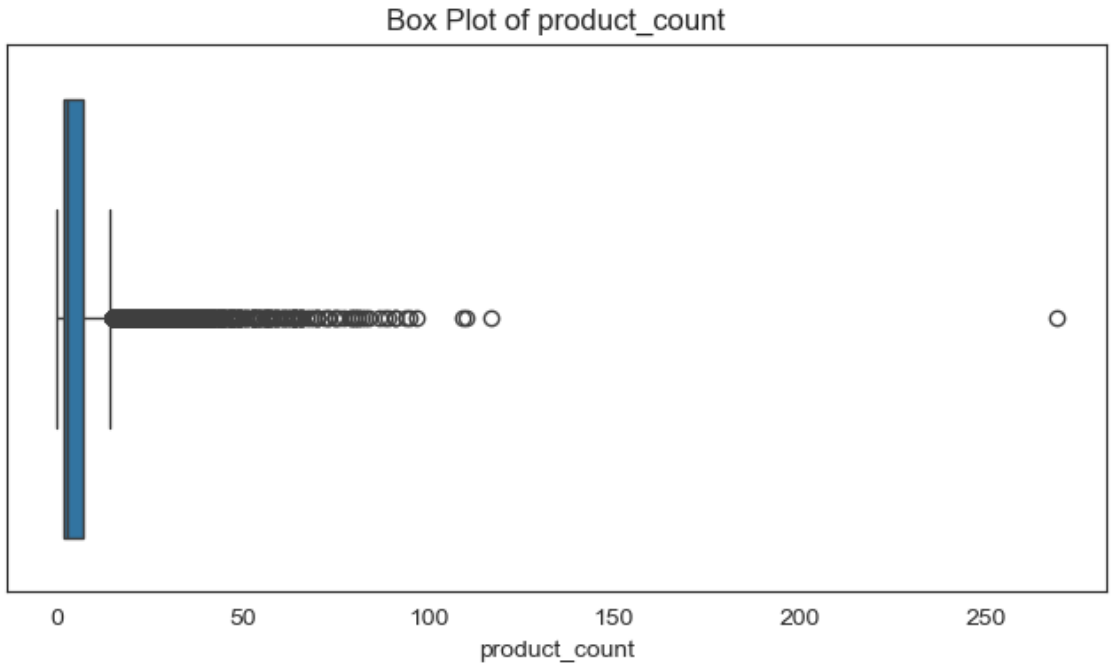

Figure 2 Box Plot of product_ount

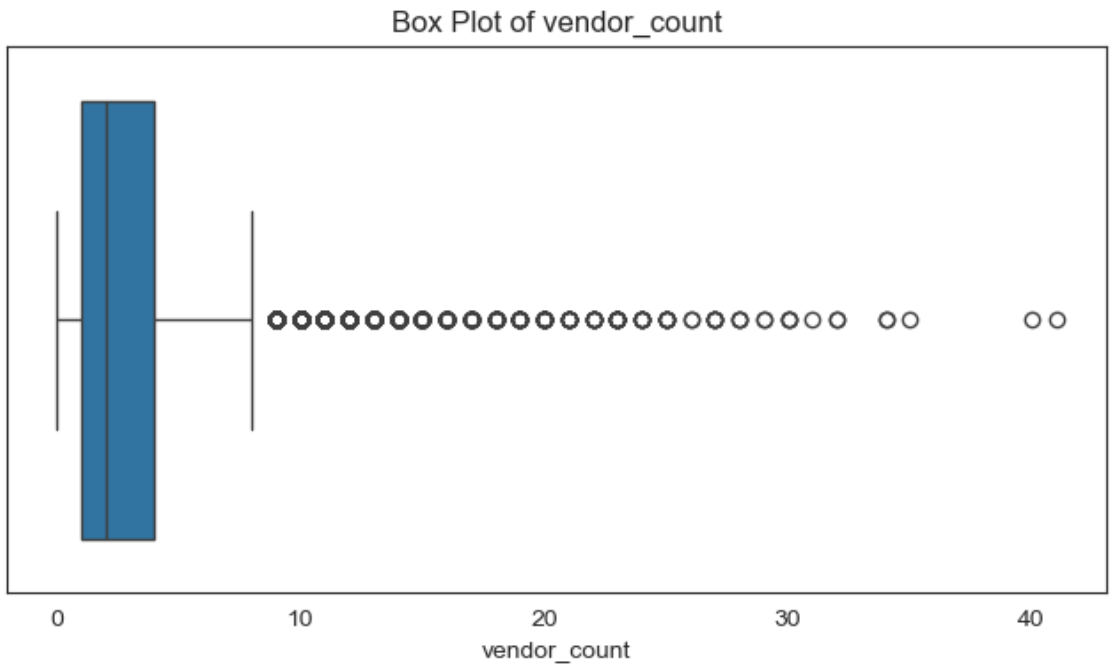

Figure 3 Box Plot of vendor_count

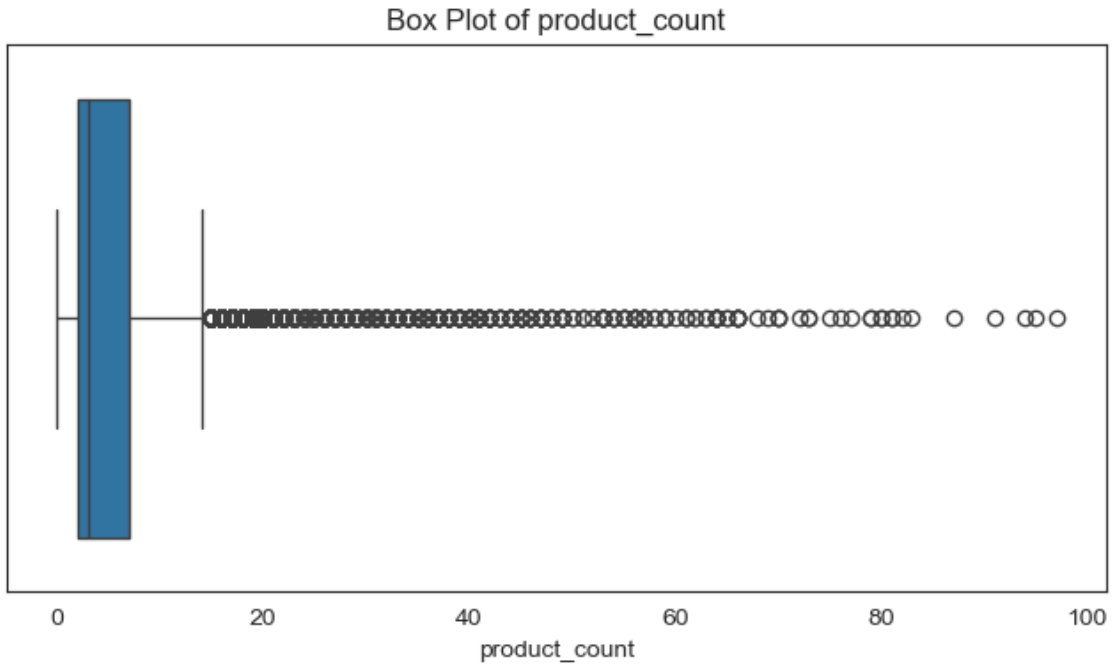

Figure 4 Box Plot of product_count after outliers treatment

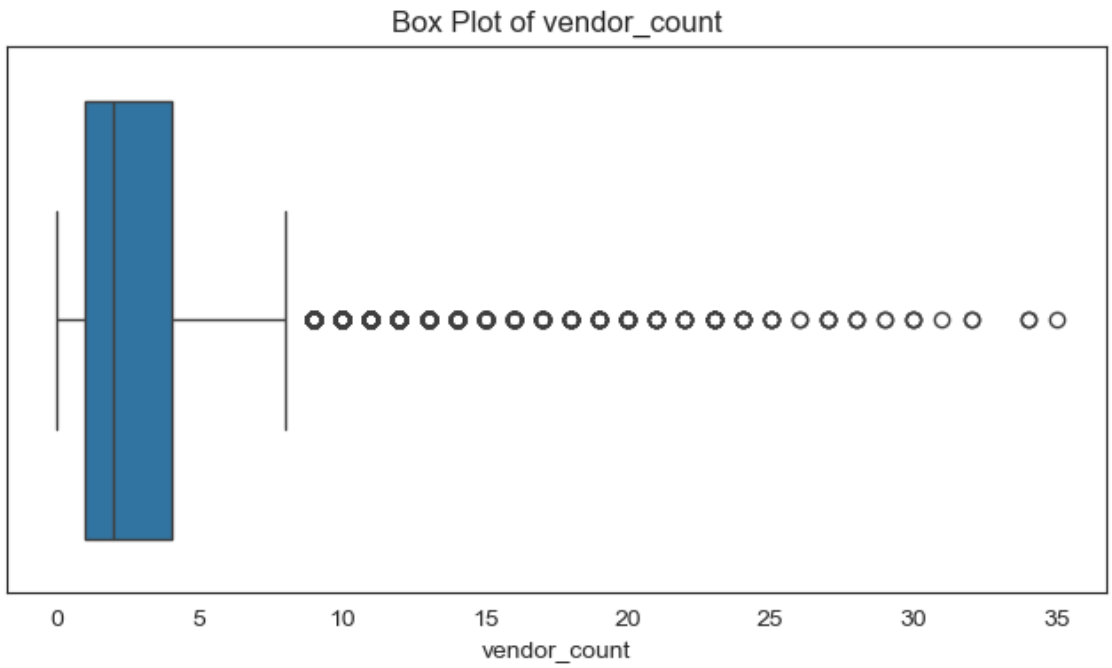

Figure 5 Box Plot of vendor_count affer Outliers treatment

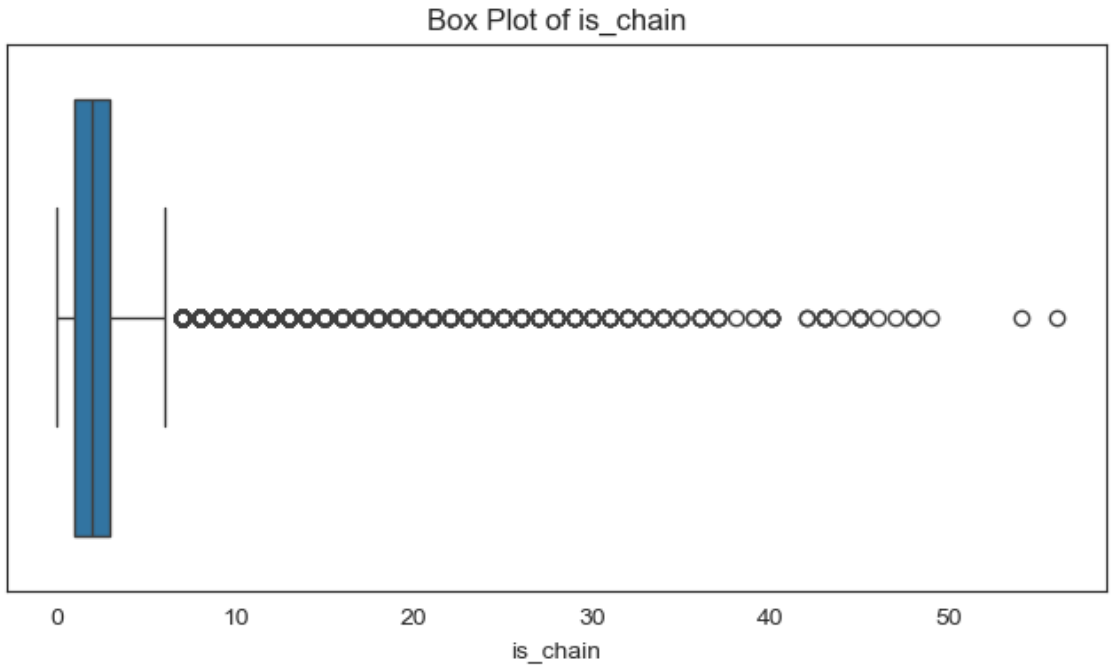

Figure 6 Box Plot of is_chain after Outliers treatment

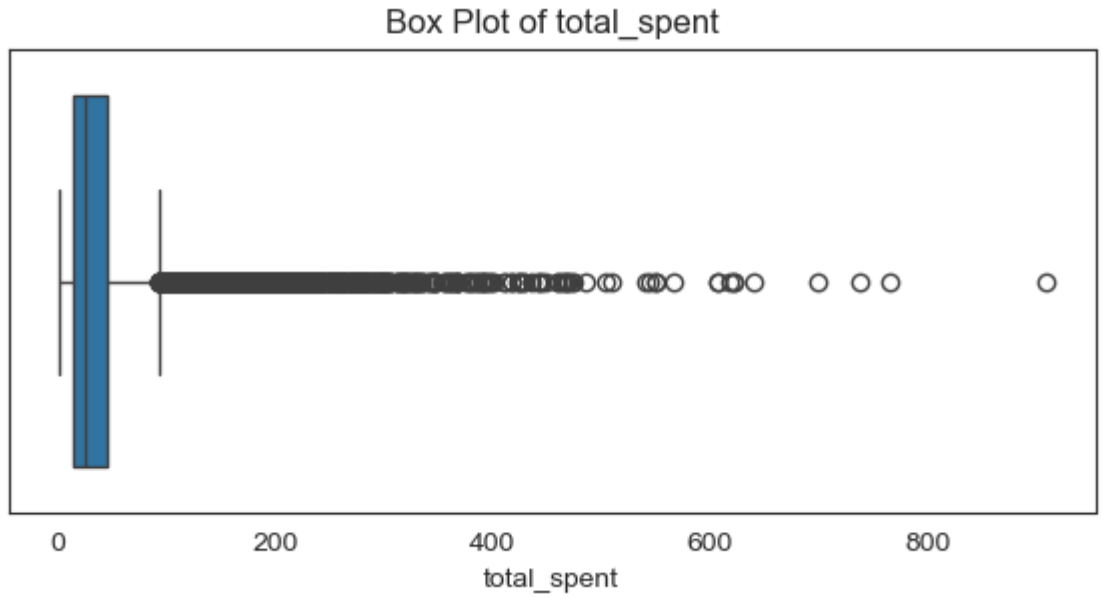

Figure 7 Box Plot of total_spent

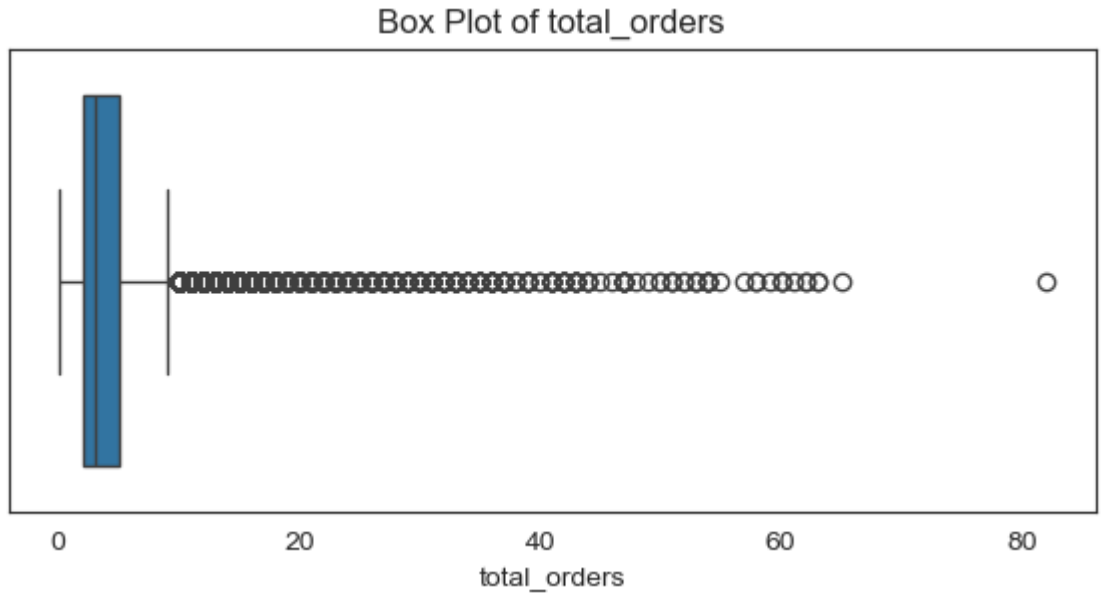

Figure 8 Box Plot of total_orders

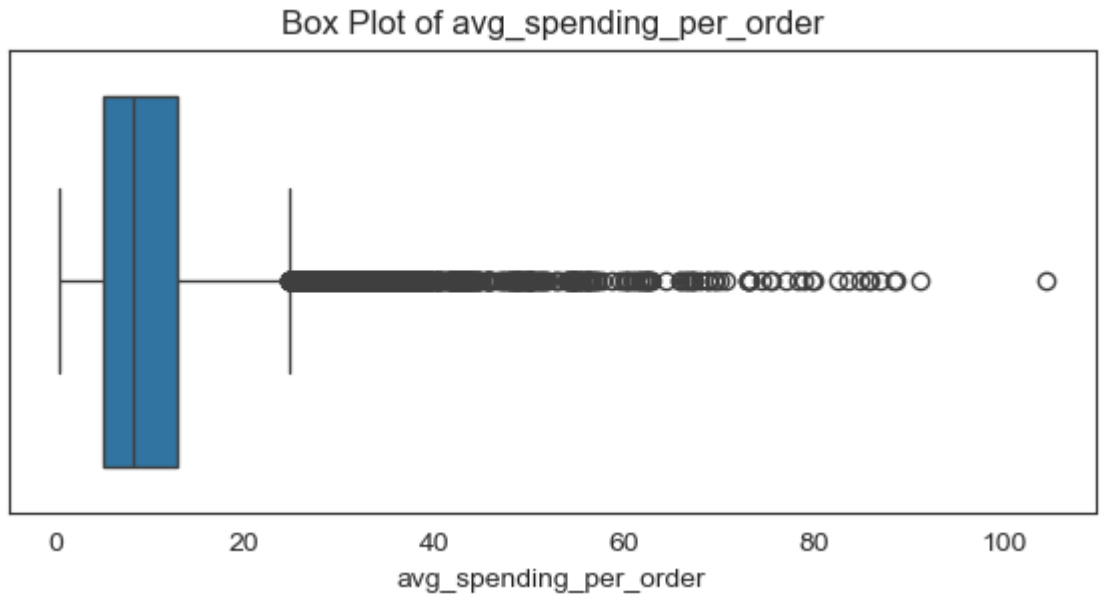

Figure 9 Box Plot of avg_spending_per_order

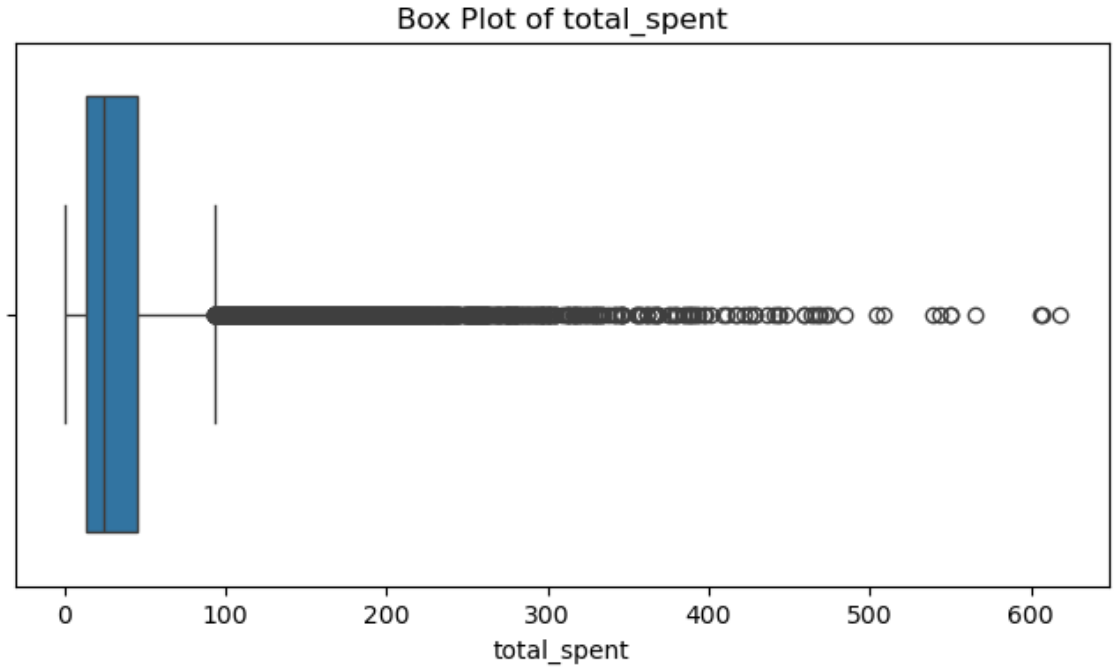

Figure 10 Box Plot of total_spent after Oultiers treatment

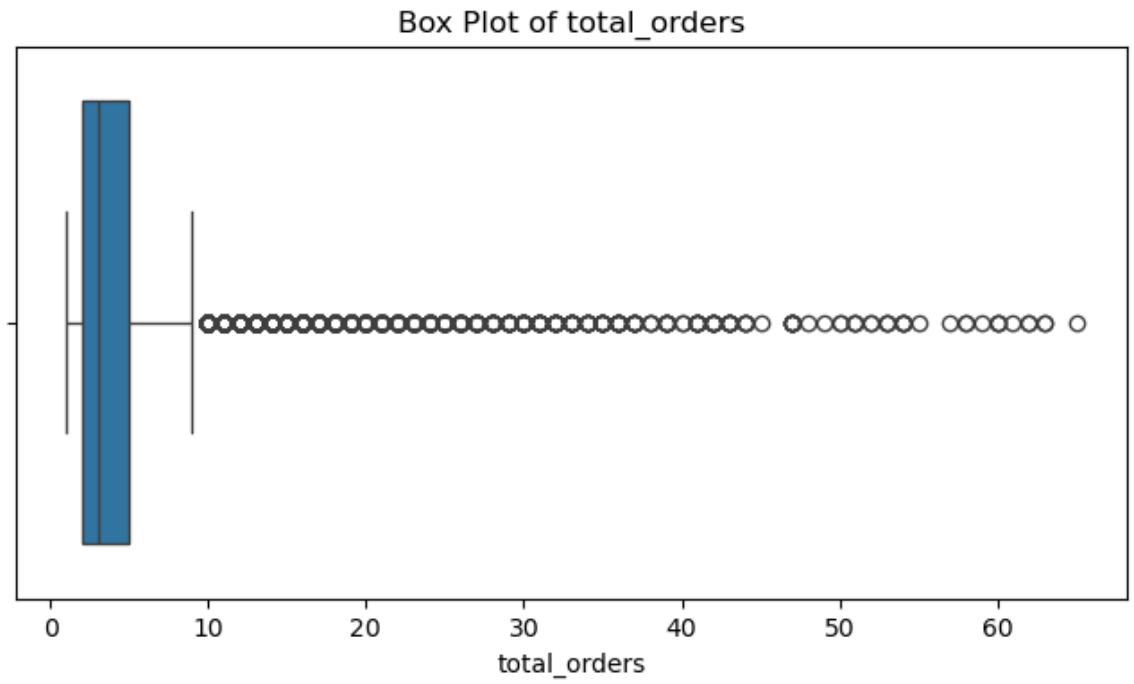

Figure 11 Box Plot of total_orders after Outliers treatment

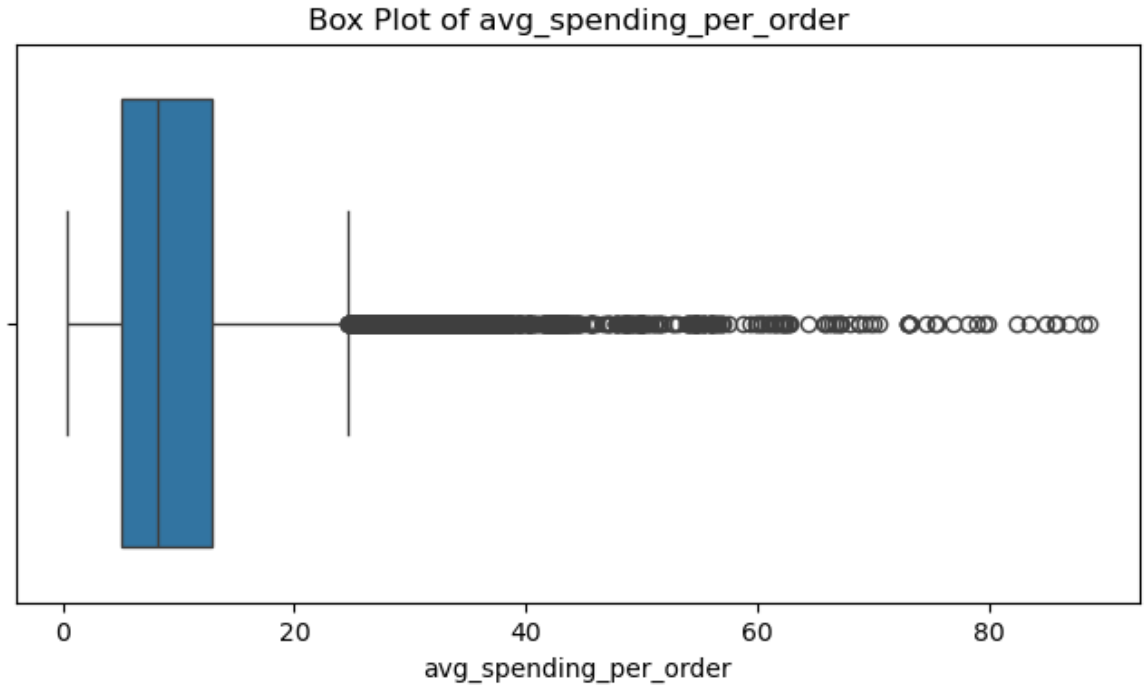

Figure 12 Box Plot of avg_spending_per_order after Outliers treatment

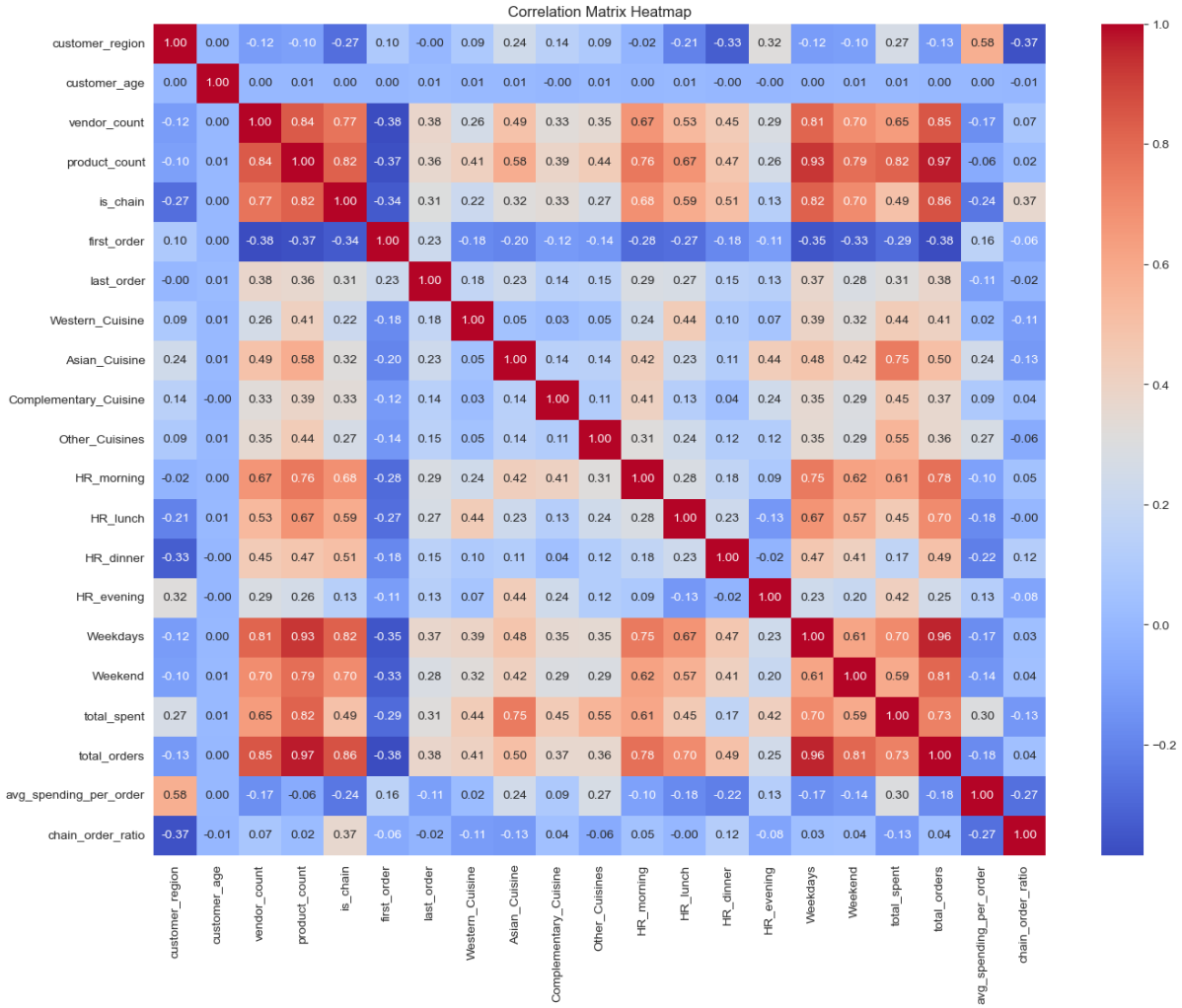

Figure 13 Correlation Matrix Heatmap

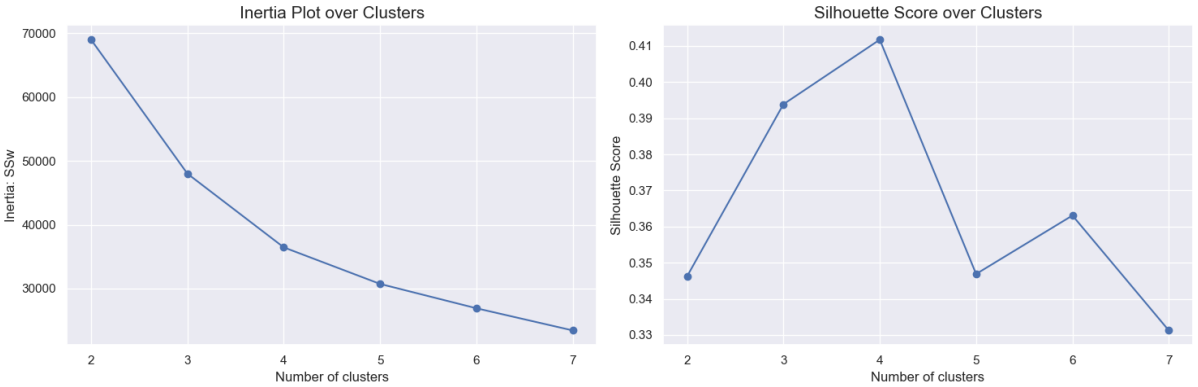

Figure 14 Inertia and Average Silhouette Plots Value perspective

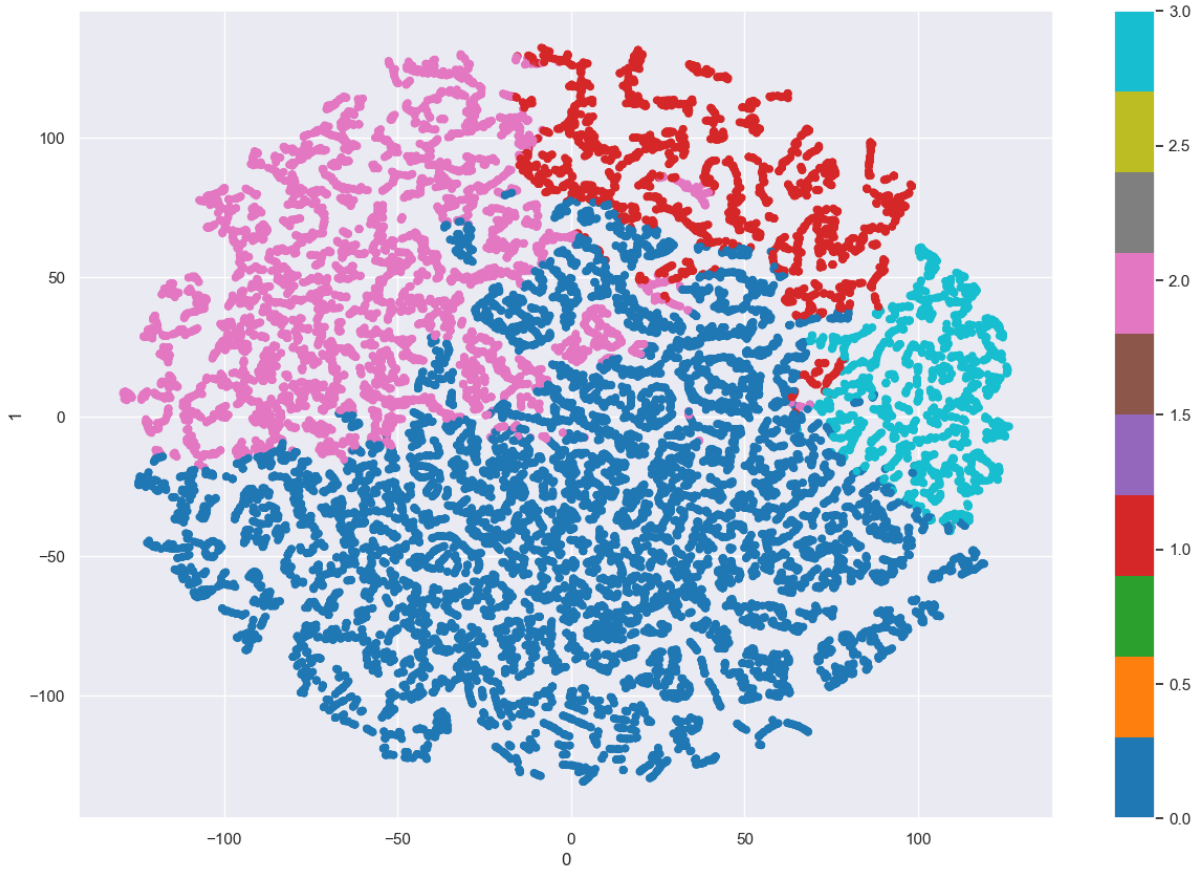

Figure 15 TSNE for value perspective

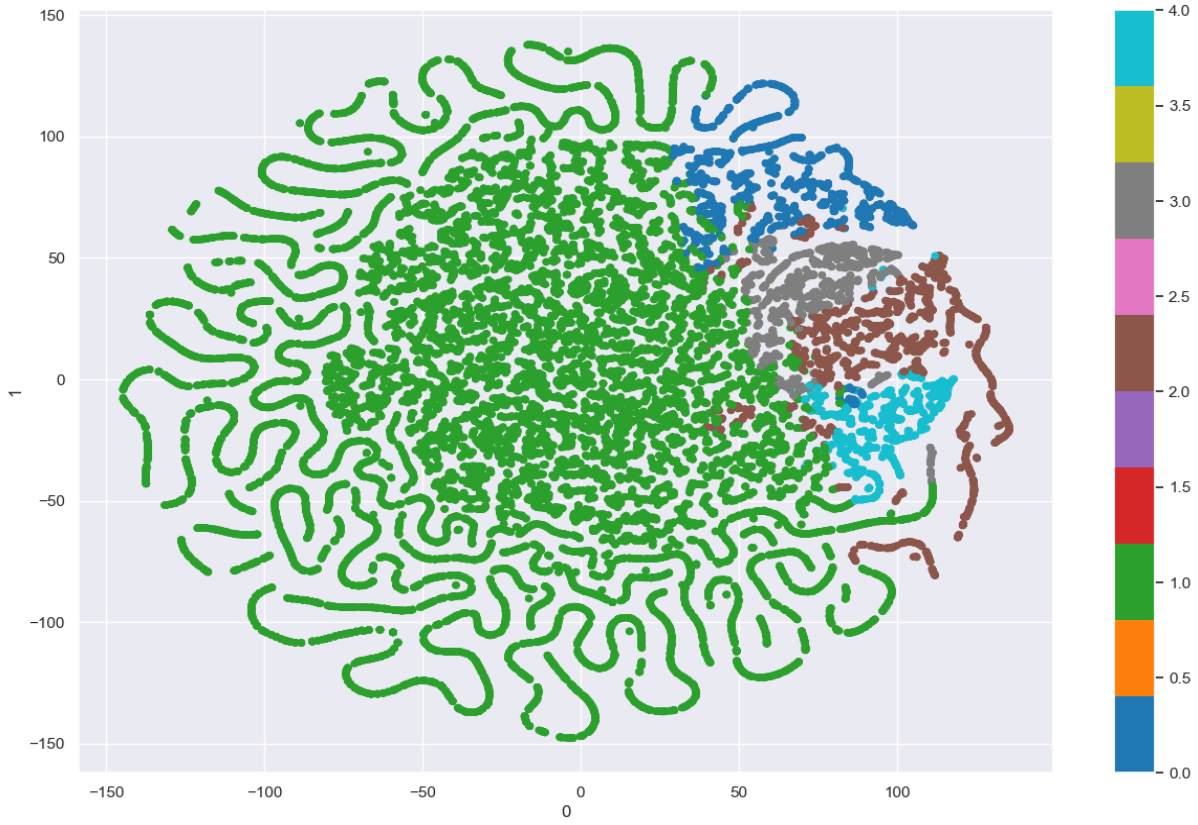

Figure 16 TSNE for preferences perspective

Figure 17 TSNE for preferences perspective

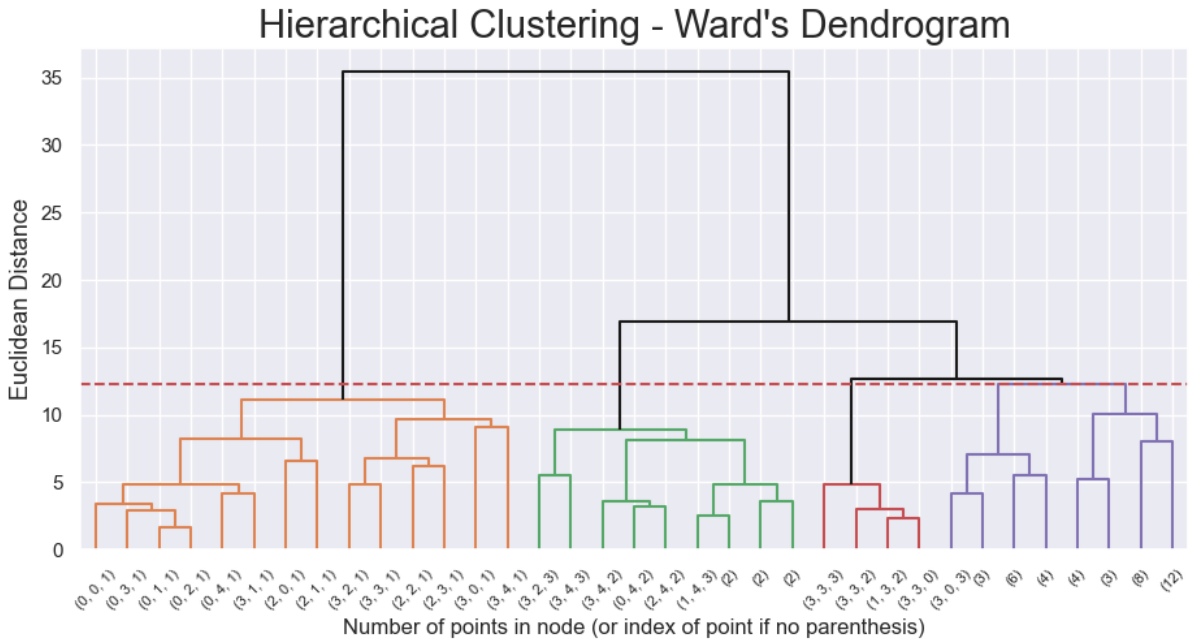

Figure 18 Dendrogram merged cluster Kmeans with HC

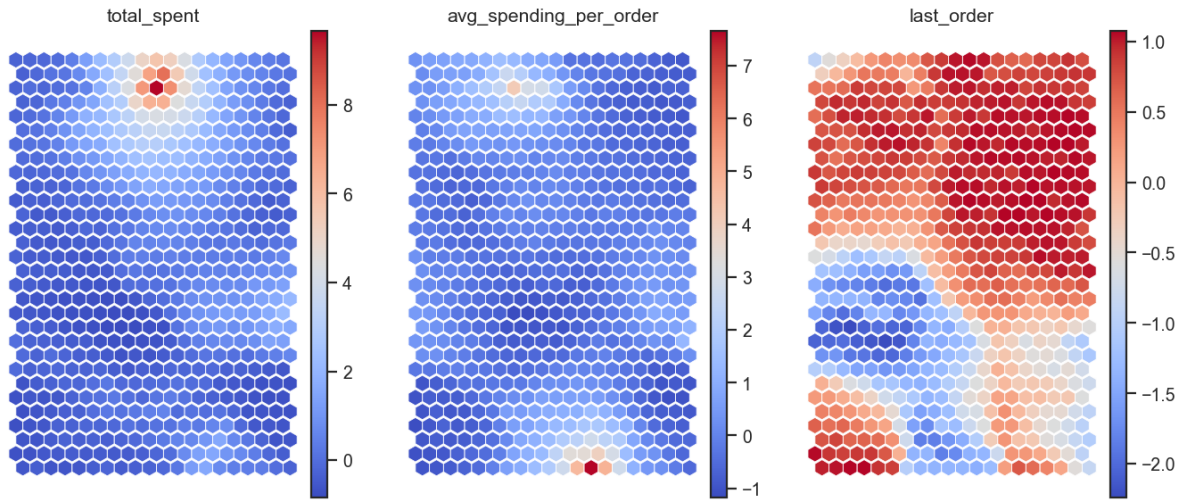

Figure 19 Hexagons SOM value perspective

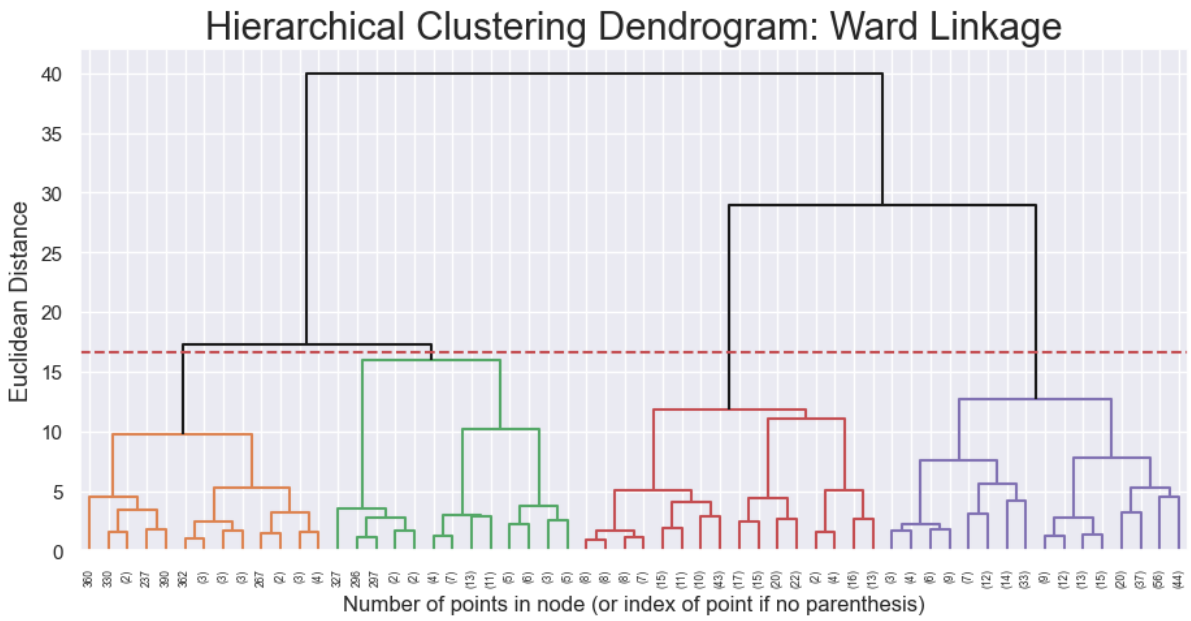

Figure 20 Dendogram HC Value

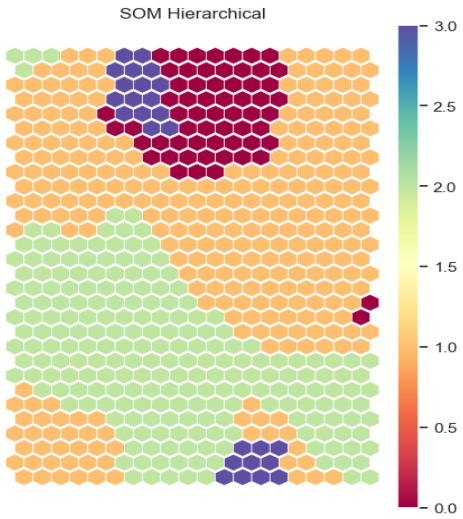

Figure 21 Hexagons SOM value perspective

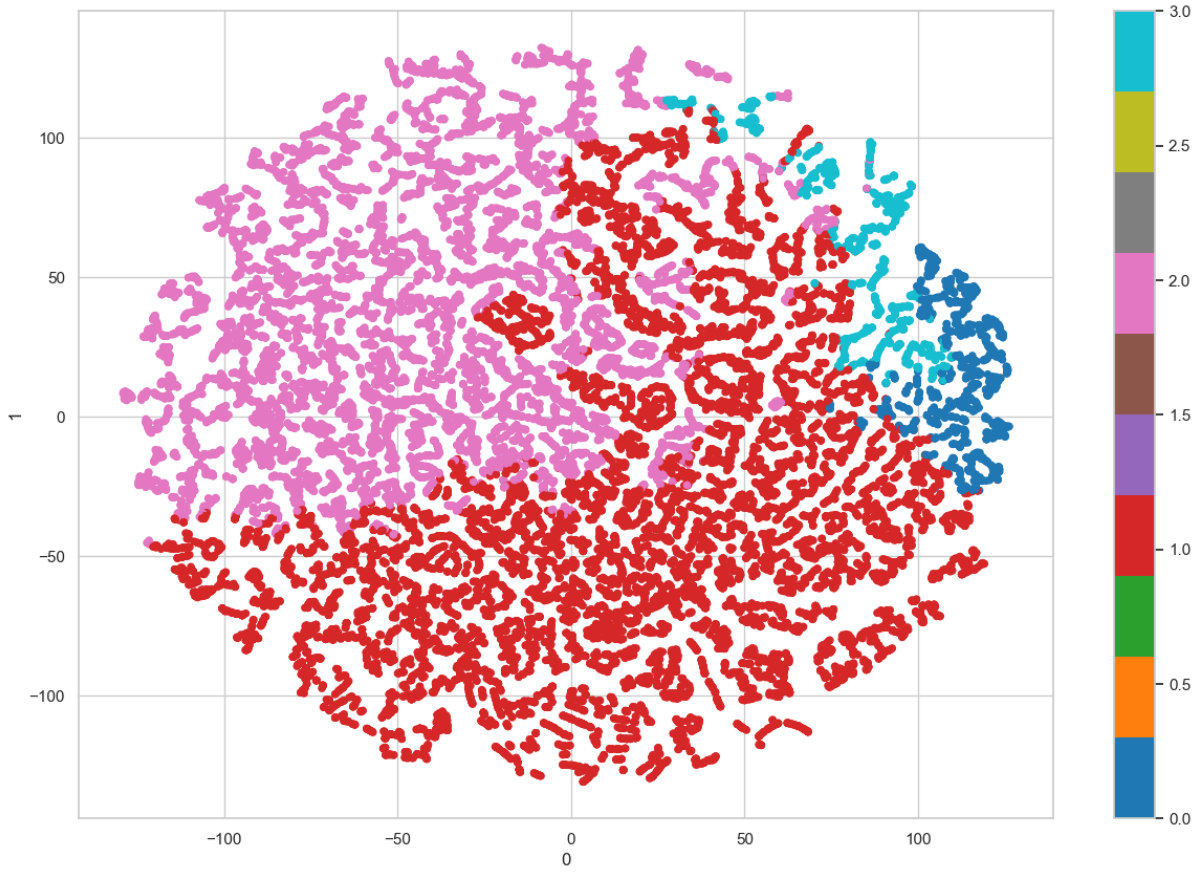

Figure 21 TSNE SOM HIERARCHICAL

value_label	total_spent	avg_spending_per_order	last_order
0	3.388332	0.172513	0.883887
1	-0.032975	-0.166616	0.657293
2	-0.418451	-0.055672	-0.993191
3	1.377297	2.985305	0.231982

Figure 22 SOM Labels for value perspective

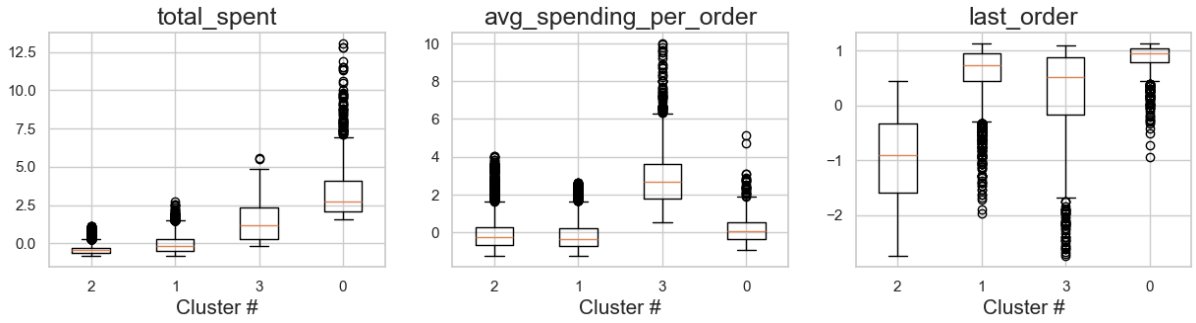

Figure 23 BOX PLOTS each Label for value perspective

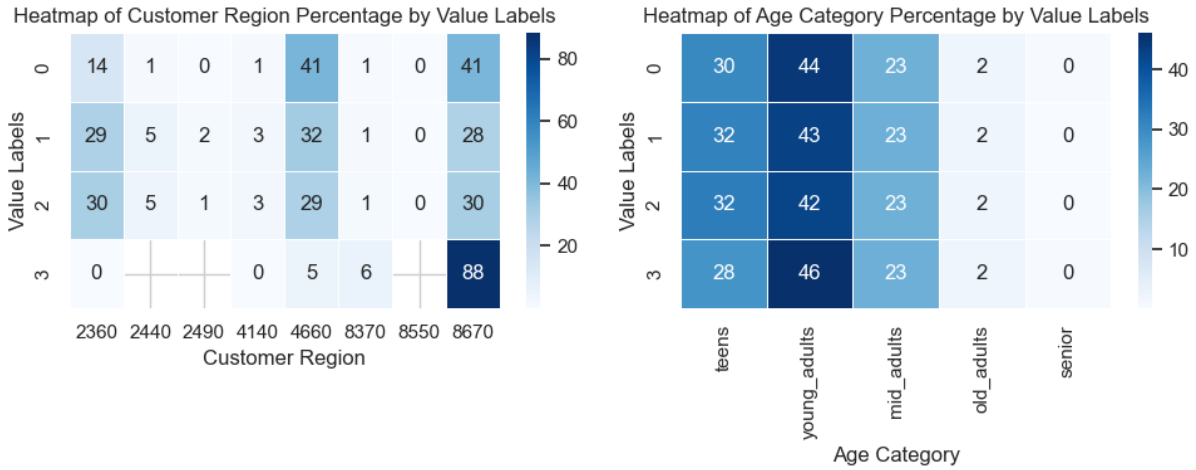

Figure 24 Association region with each Label for value perspective

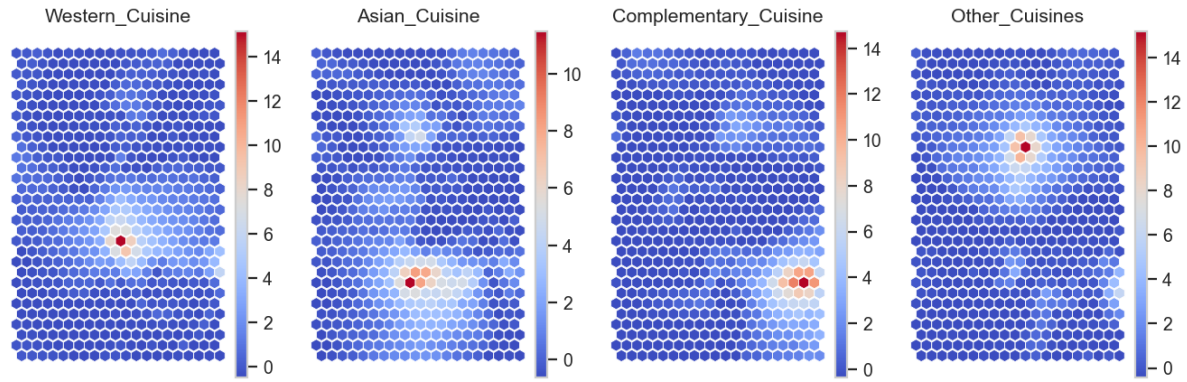

Figure 25 Hexagons SOM preferences perspective

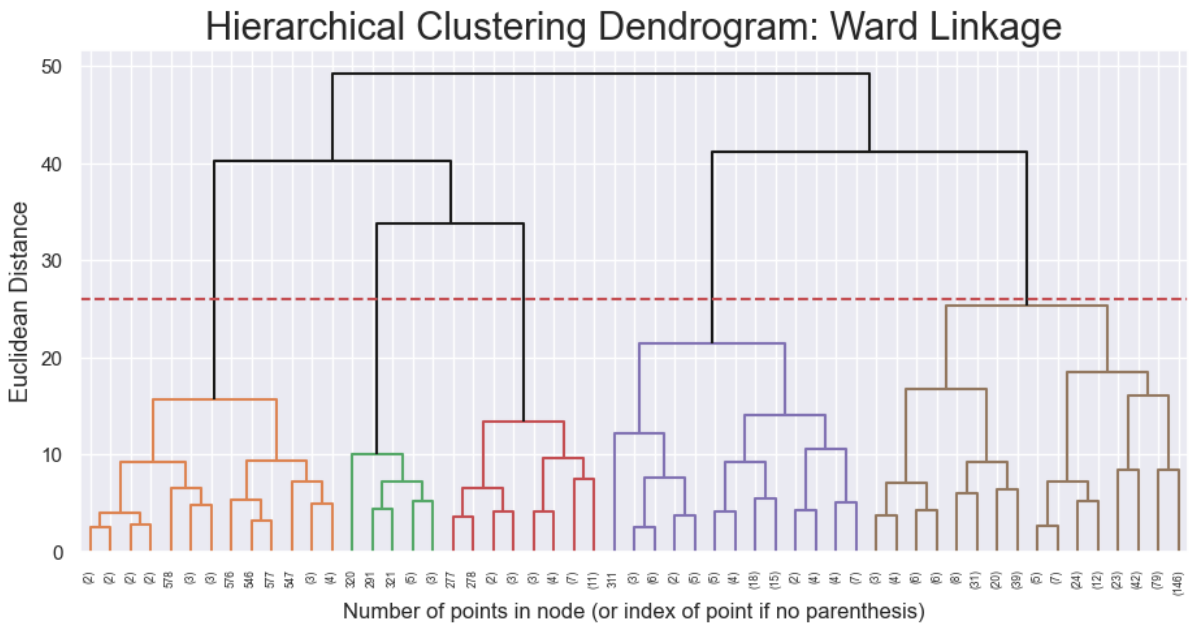

Figure 26 Dendogram HC preferences perspective

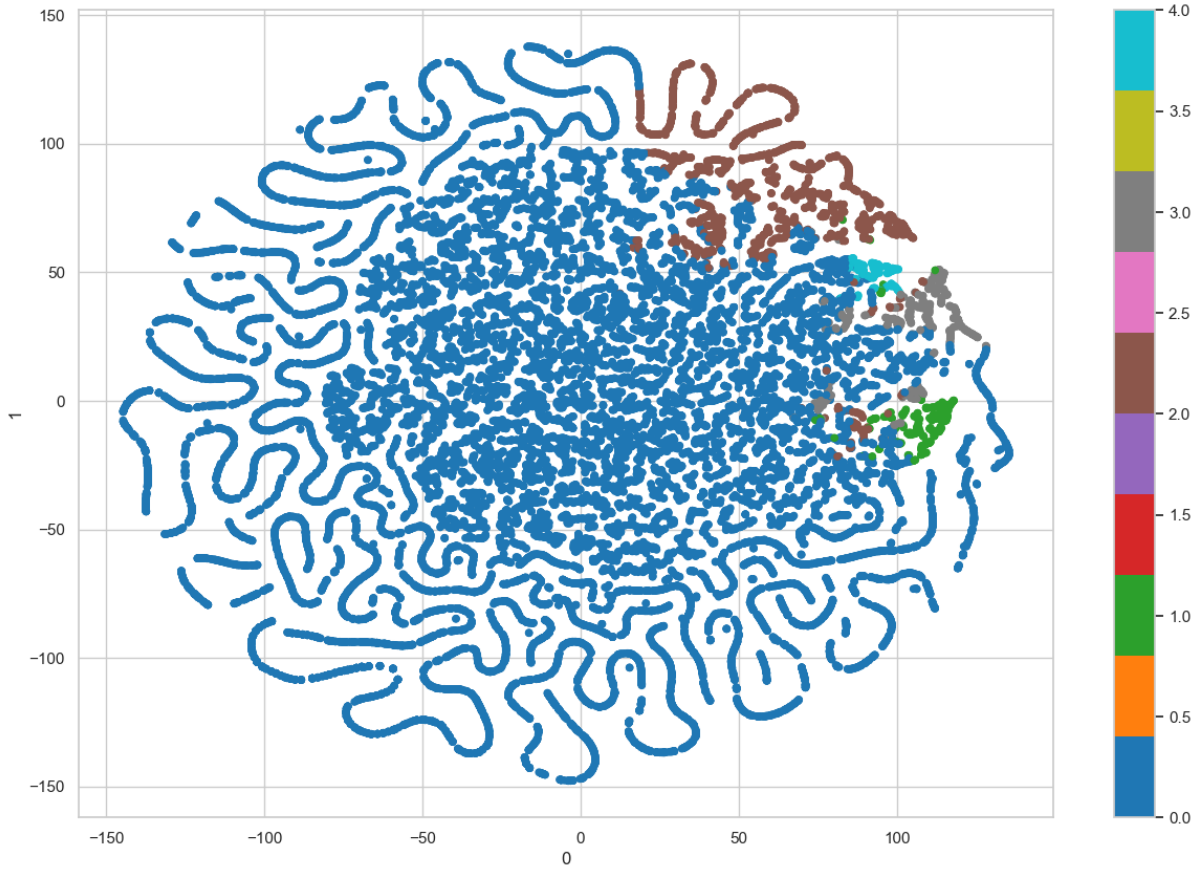

Figure 27 TSNE SOM HIERARCHICAL preferences

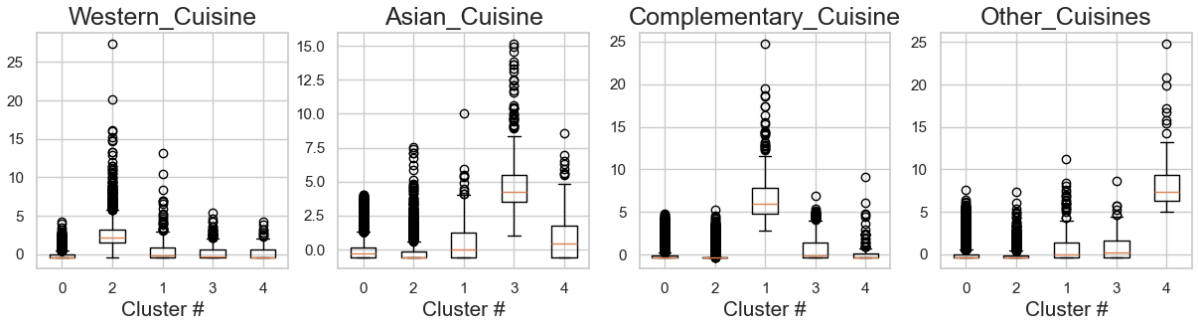

Figure 28 BOX PLOTS each Label for preferences perspective

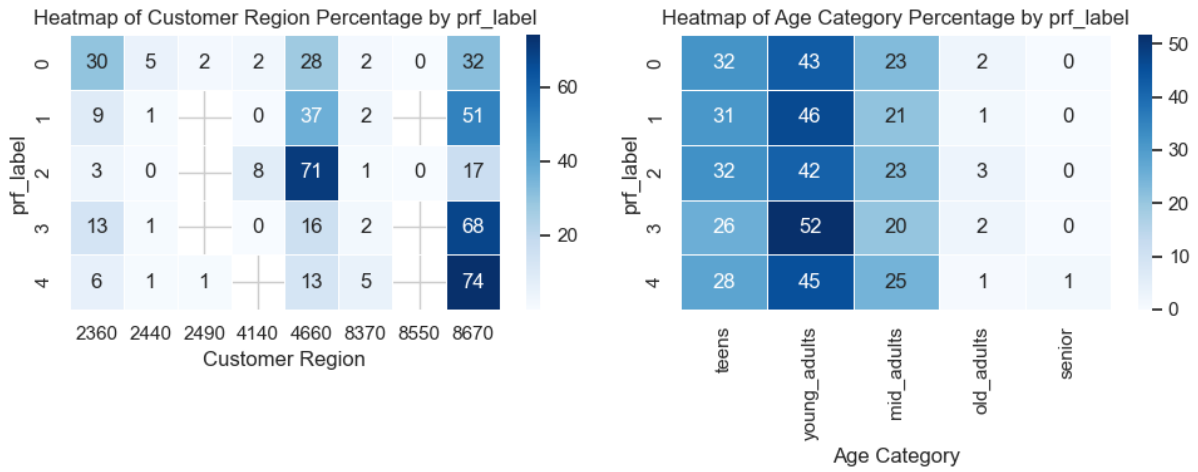

Figure 29 Association region with each Label for preferences perspective

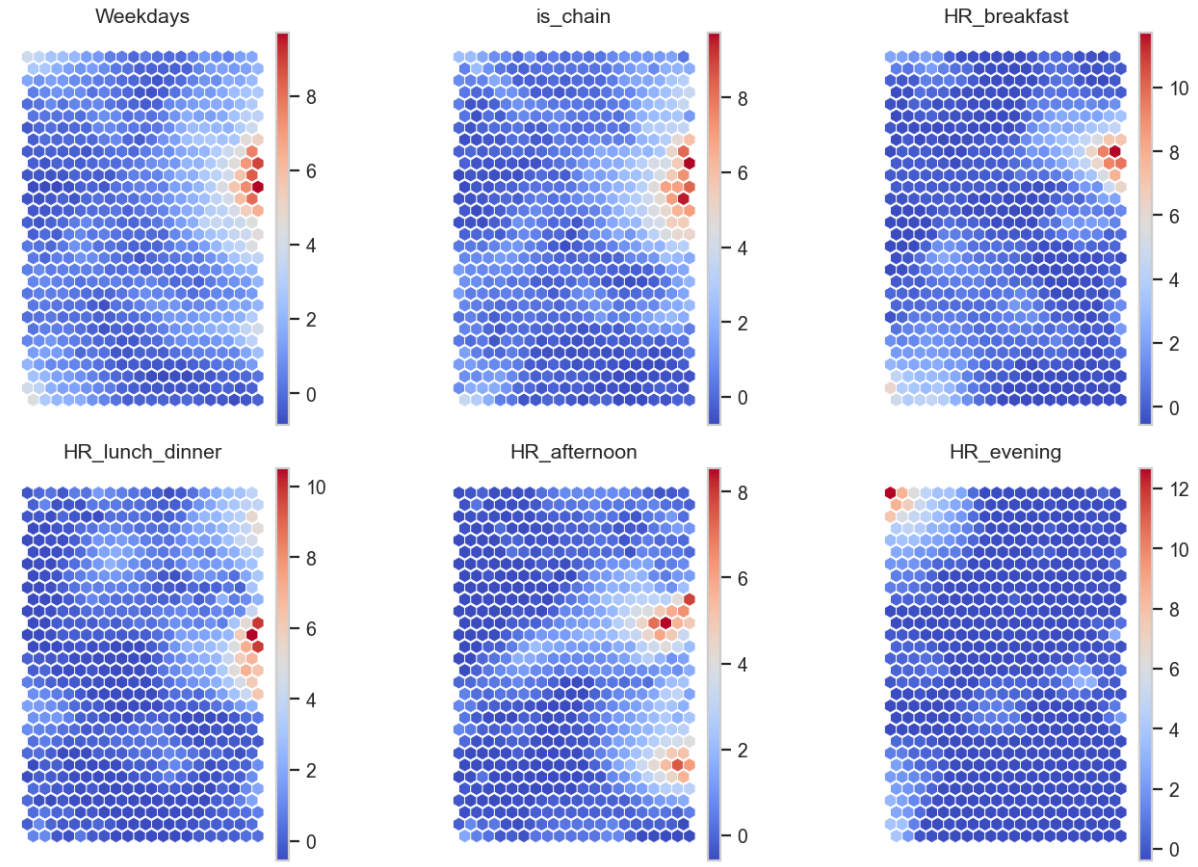

Figure 31 Hexagons SOM behavior perspective

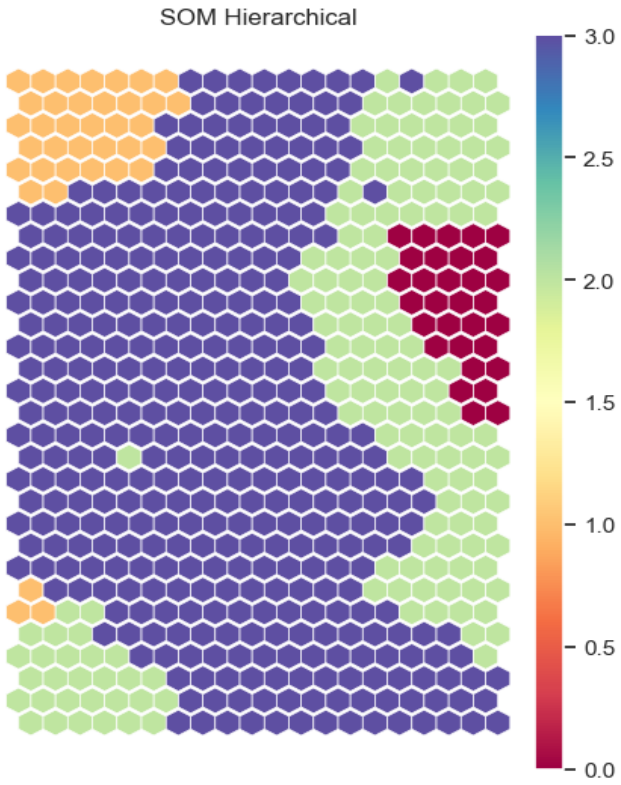

Figure 32 Hexagons SOM behavior perspective

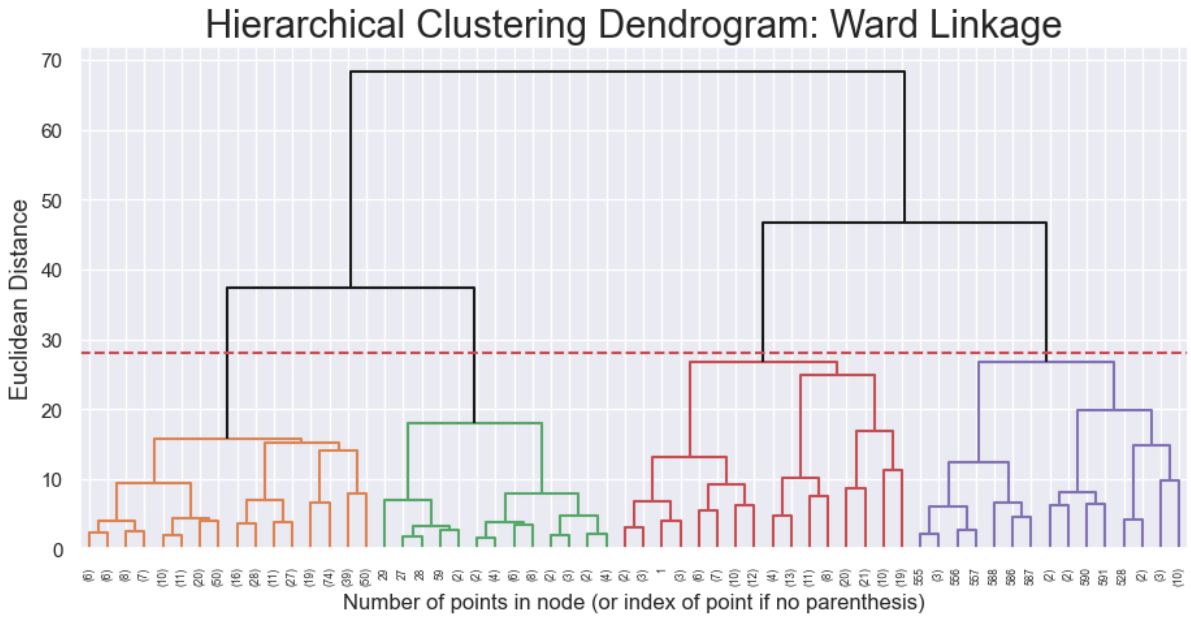

Figure 33 Dendogram HC behavior perspective

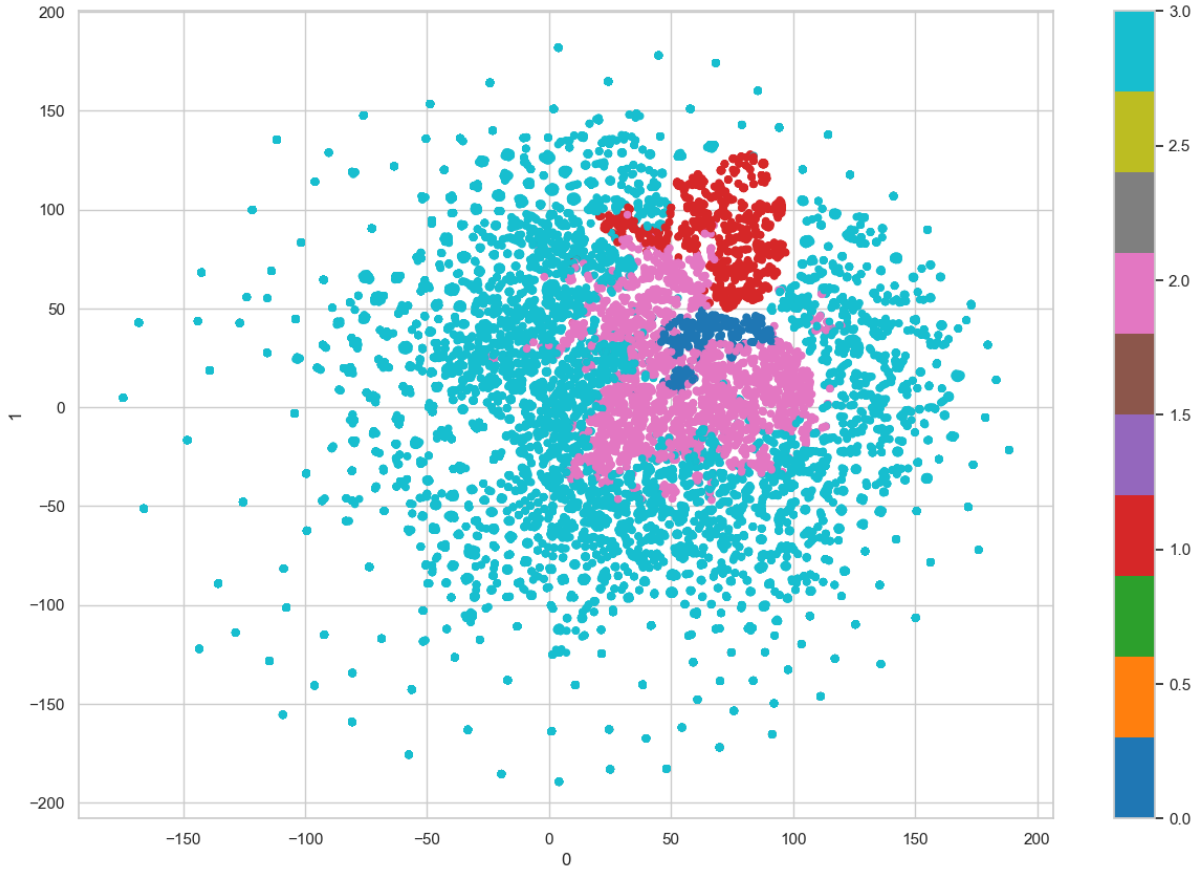

Figure 34 TSNE SOM behavior preferences

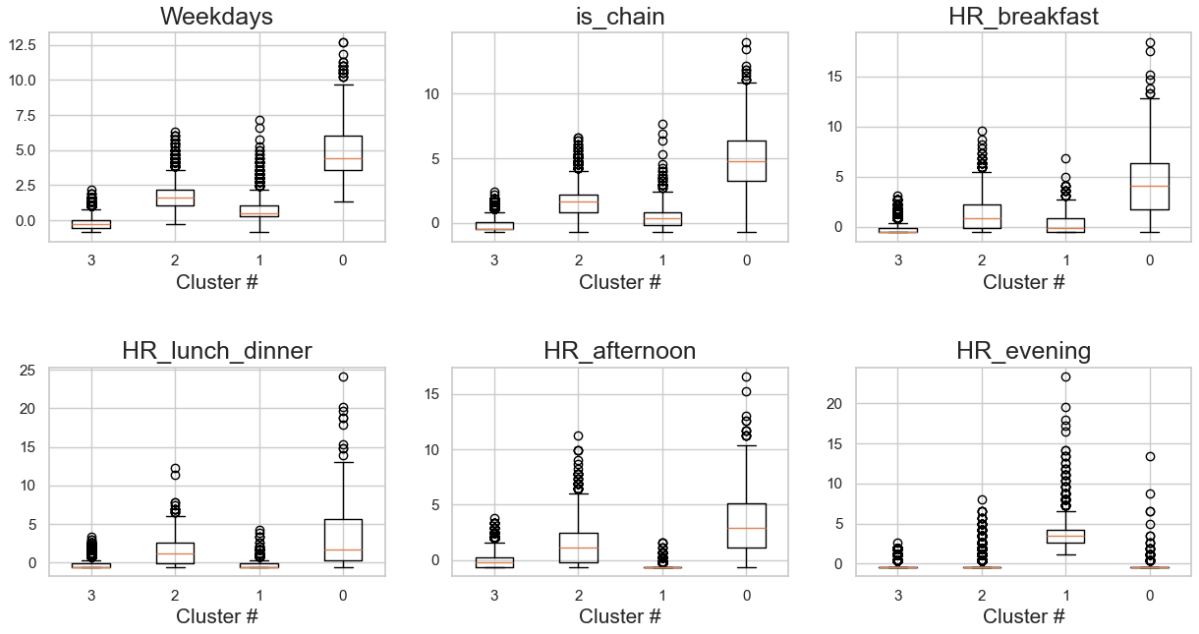

Figure 35 BOX PLOTS each Label for behavior perspective

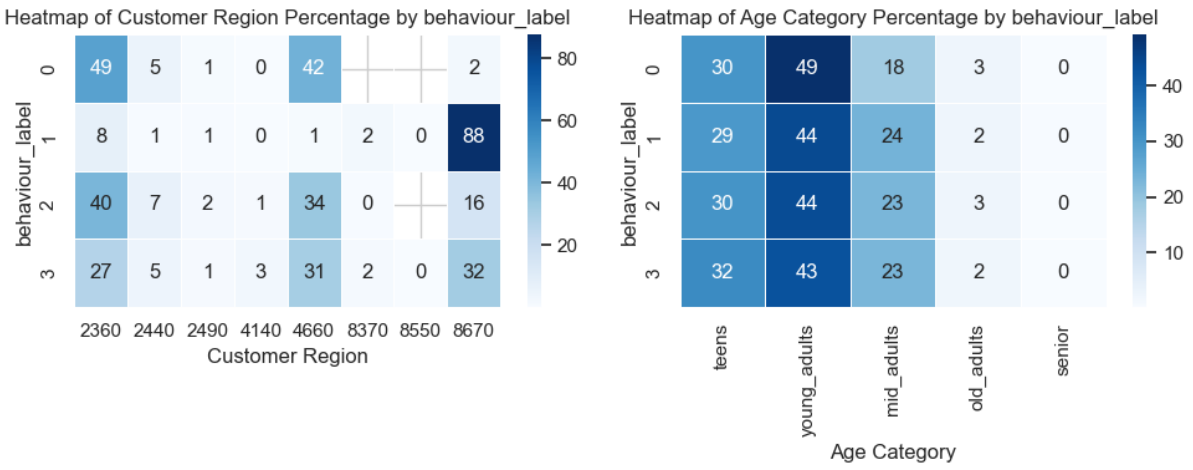

Figure 36 Association region with each Label for behavior perspective

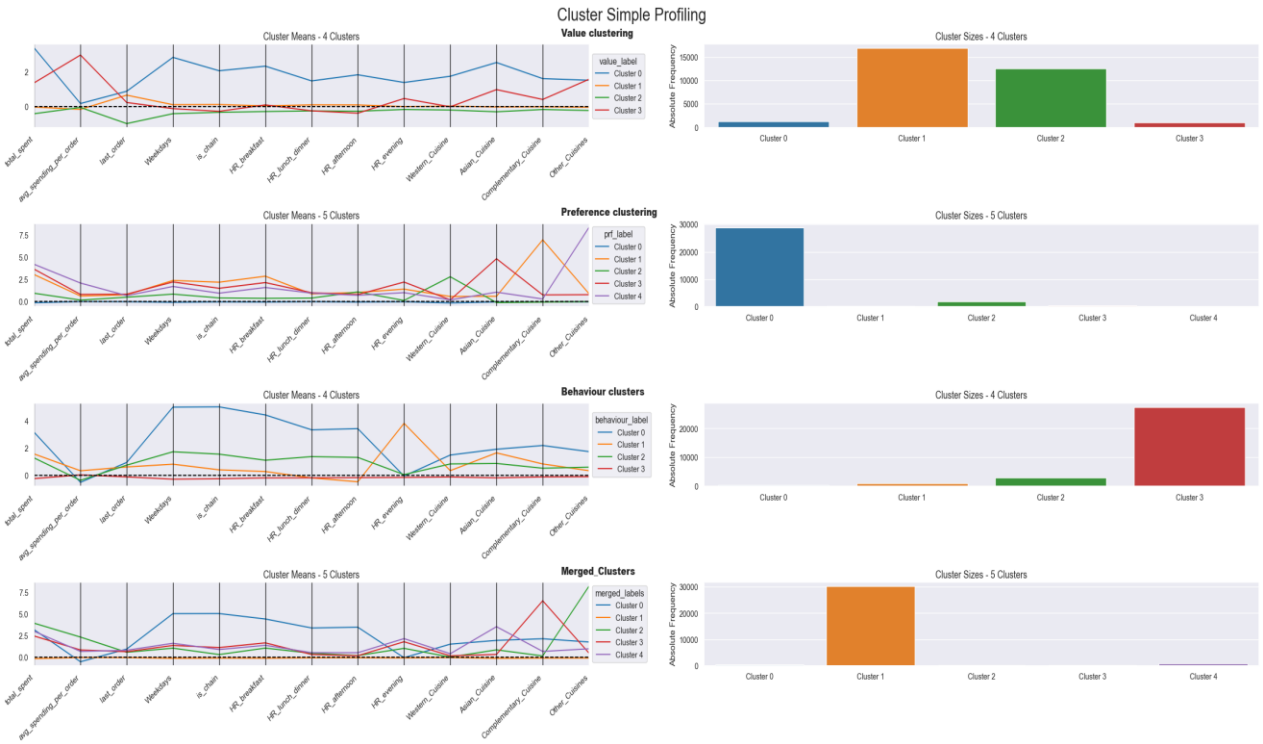

Figure 38 Cluster Analysis SOM with HC

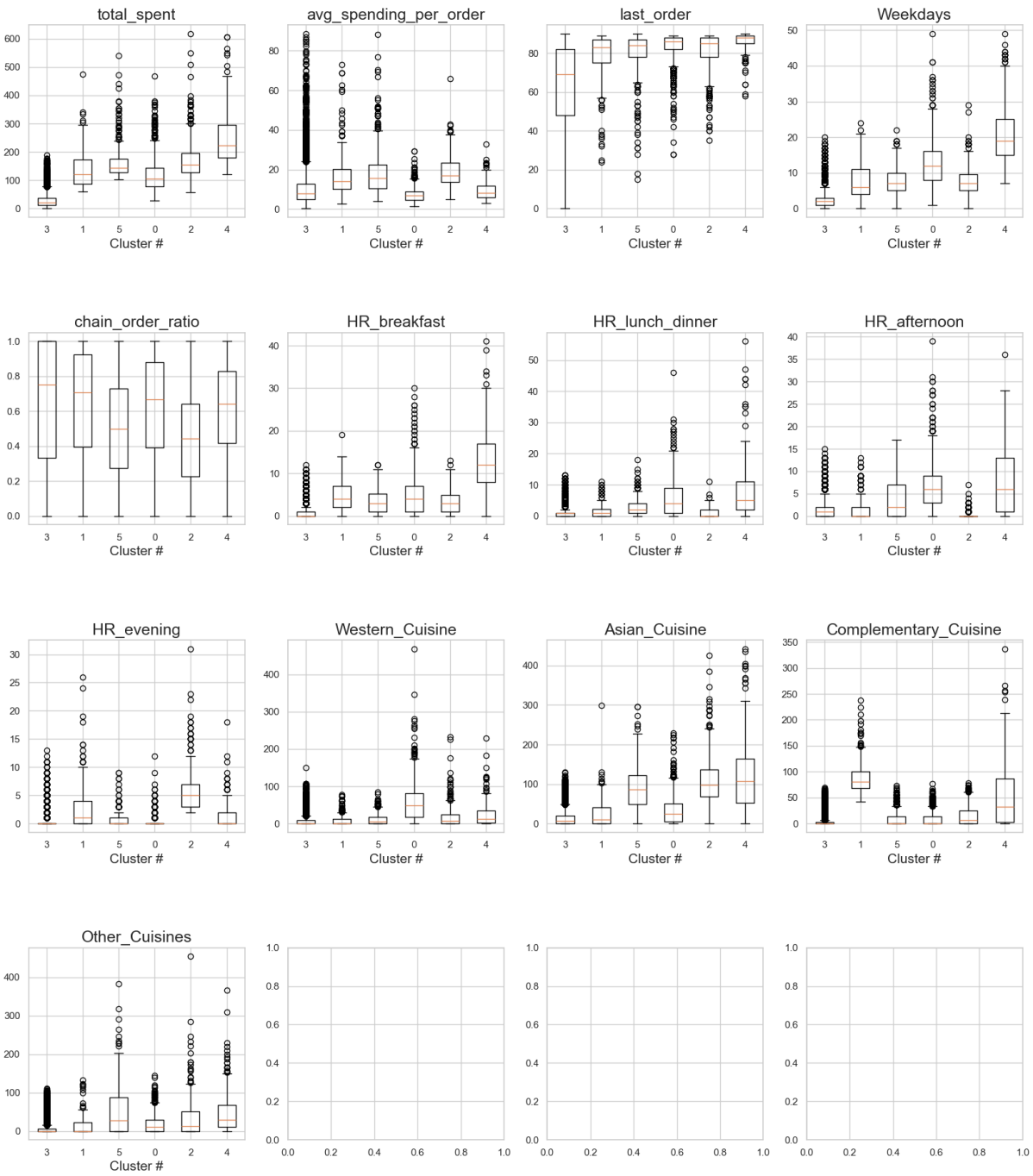

Figure 39 Boxplots metric features after SOM

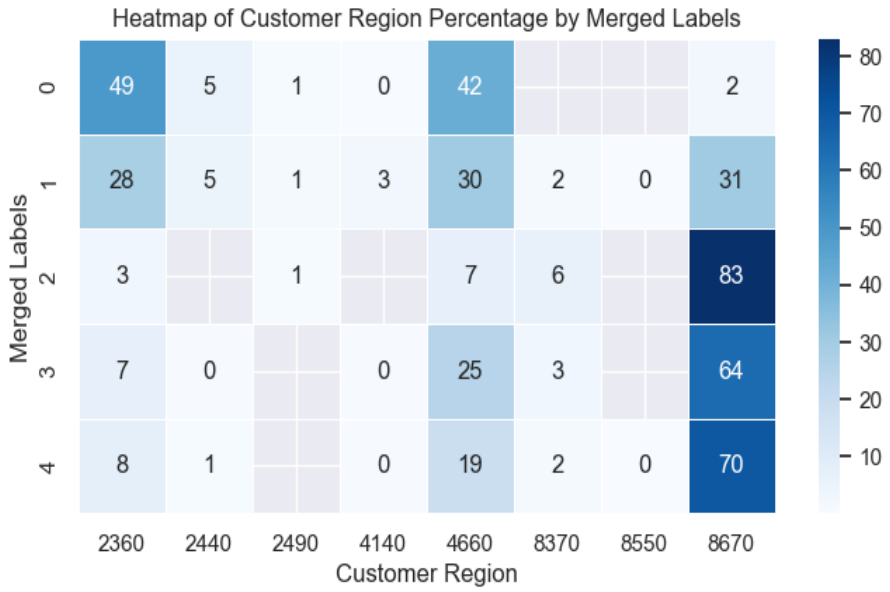

Fig 40 – Final merged clusters association with region

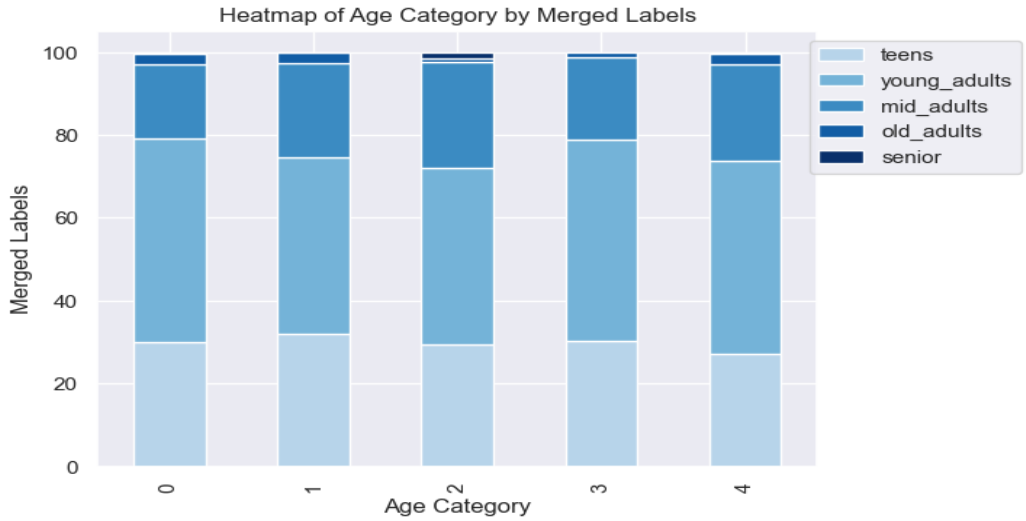

Fig 41 – Final merged clusters association with age category

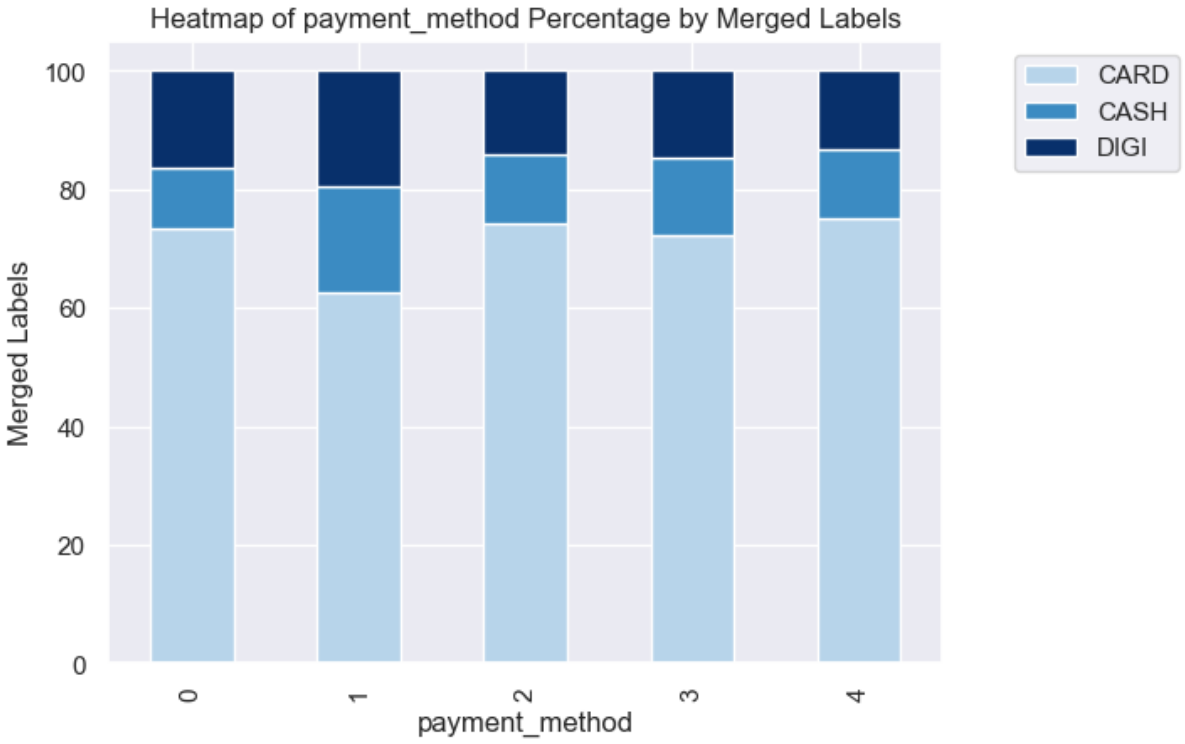

Fig 42 – Final merged clusters association with payment method

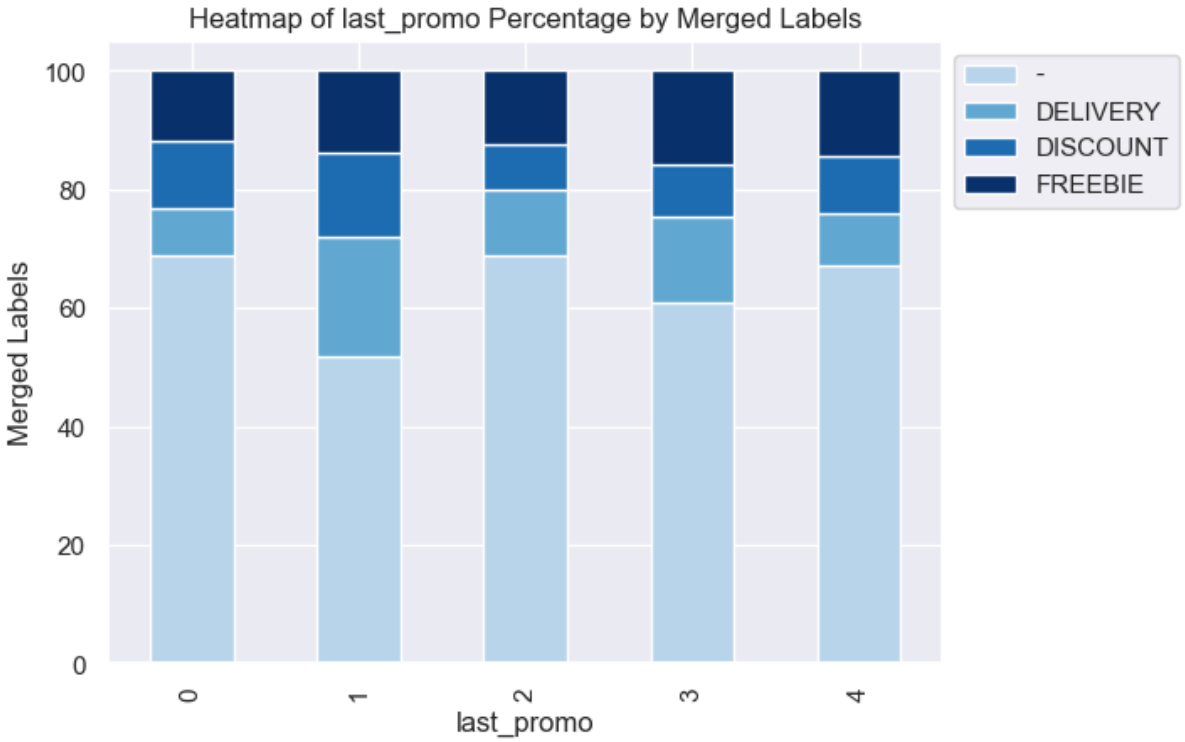

Fig 43 – Final merged clusters association with last promo