DATA MINING PROJECT

Master in Data Science and Advanced Analytics

NOVA Information Management School

Universidade Nova de Lisboa

Exploratory Analysis of ABCDEats Customer Data

Group 41

Elcano Gaspar, 20241021

Gaspar Pereira, 20230984

Pedro Costa, 20222121

Rita Wang, 20240551

Fall/Spring Semester 2024-2025

TABLE OF CONTENTS

1. INTRODUCTION

Lilien and Rangaswamy (2003, as cited in Dolnicar, Grün, & Leisch, 2018) market segmentation is essential to a successful marketing strategy, with top firms using segmented consumer groups to align their approaches and maximize impact. For this project, ABCDEats Inc., a food delivery service, we apply a data-driven segmentation approach to analyze three months of customer data across different regions. The analysis will identify key consumer segments (multiple segmentation perspectives) and help ABCDEats craft a tailored strategy that aligns with unique consumer preferences.

2. DATA EXPLORATION

2.1. Data Overview

The ABCDEats Inc. dataset contains 31,888 rows each consisting of a client from the app, and 56 features, with detailed information about customer activity (table 1). We found 13 duplicates (customers with the same id, and same buying patterns) that were removed from the dataset. We observed that first_order, customer_age, HR_0 should be integer variables but were represented as float as they had some null values.

2.2. Univariate Analysis

Regarding the numerical features we examined their distribution using histograms and boxplots shown in figure 1. The overview of the numerical features statistics is presented in table 2. For age, half of customers are young adults aged 23 to 31 years old. Customer age range from 16 to 80 years old. From age 45, there are fewer consumers. Most customers purchased from 1 to 4 different vendors, at least 2 to 7 different products.

Regarding the "is_chain" feature, the metadata seems misleading as it appeared to be a Boolean variable, but it is a numerical variable with values from 0 to 83. We interpreted this variable as the number of orders made from chain companies. We can observe that these 2 variables have a right-skewed distribution (figure 1), with a minority of customers making orders from many different vendors and chain companies, with multiple products.

The day with the most customers making their first order during the collection of the dataset was the first day (1542) (figure 2). Since the data collection began after the service was already operational, this higher number of first orders, in the beginning, reflects the existence of an already established user base. The number of first orders shows a downward trend, reaching 121 first orders by the 90th day. It’s likely that this pattern occurs because at first both existing users of the service and new customers are making their first orders. Over time, the number of first orders decreases, and eventually plateaus, as most of the active customers have already made their first orders and the numbers mainly reflect new users joining the service. The histogram reveals periodic fluctuations, with notable peaks about every seven days, indicating weekly patterns in customer behavior leading to new orders. This may be due to weekends or use of promotional codes. For the left-skewed distribution of the last_order, on the 90th day, there are 1711 last orders, which shows that the number of daily active users per day remained fairly consistent over the three-month period. Like first_order, the last_order distribution shows a weekly pattern, where user activity gradually increases and peaks roughly every seven days, followed by a slight dip before rising again. We observed that both variables include data for a 91st day, with one first order and 32 last orders. This anomaly could be from incomplete data collection on this day.

Analyzing total spending on different types of cuisines, there is a strong preference for Asian cuisine, which accounts for approximately 26% of the total spending value (figure 3). To better understand the spending patterns and reduce outlier influence, as for each type there are a few high spenders, we applied a (natural) log transformation to cuisine spending variables. On average, customers spend 1.95 monetary value on Asian food, and second comes American cuisine accounting for around 12.74% of the total and an average of 1.24. In contrast, only 1.86% is spent on Noodles, with an average of just 0.15 spent per customer. Low average spending is explained by the fact that most cuisines, except for Asian and American (62.68% and 66.15%, respectively), have 75% or more customers that have never ordered from them. This suggests that customers tend to limit themselves to a small range of cuisines.

For total orders by day-of-week, DOW_0 to DOW_6, distribution appears similar, with only small volume variations. The demand is stable during the week rising slightly towards the end of the week on Thursday, Friday, and Saturday while Sunday has the lowest volume (figure 4). This behaviors are likely due to weekend shopping or leisure. For total orders by each hour of the day, HR_0 to HR_23, there are two peaks in the distribution: morning (10H-12H), and evening (16H-18H). This behavior is often associated to main meal times. A third peak at 3H-4H was also observed, probably related to nighttime activities (figure 5).

For categorical features, we analyzed the distribution shown in figure 6, with a summary of the distribution in table 3. Most customers are from the region "8670", followed by "4660" and "2360". Smaller portions come from other regions like: "2440", "4140", "8370", "2490", and "8550". There is also 442 customers with unidentified regions. The number of customers who used a promo in the last order is balanced (around 47%) compared to those who did not (53%). Of those who used a promo, the majority used "DELIVERY", followed by "DISCOUNT" and "FREEBIE". For payment methods, most prefer to pay by Card (around 63%), with the other two methods being quite balanced: Digital (19%) and Cash (17%).

3. DATA PREPROCESSING

3.1. Feature Engineering and Analysis

For feature engineering, we created the following features: “total_orders”, “percentage_chain”, “customer_activity”, “customer_lifetime”, “order_frequency”, “avg_time_between_orders”, “promo_used”, total spent “avg_order_value”, “favorite_cuisine”, “favorite_day”, “favorite_hour”. Table 4 in the annexes describe these new features and how they were built.

Regarding new numerical features we can see their distribution in figure 7, and an overview of their statistics in table 5. For "total_orders", the dataset has a total of 31,018 orders, with an approximate average of 10,000 orders per month. We also noticed that 138 rows of customers did not place any orders, so we dropped these. The average number of orders per customer is approximately 4, with a standard deviation of 5, showing moderate variability. For "percentage_chain" we see that most customers placed orders from chain companies (median 71%). The mode of customer lifetime is zero (23%), meaning many customers do not return to place additional orders. However, the average time between orders is 8 days, showing healthy engagement. The maximum time between orders is around 44 days, but the number of customers that exhibit such extended gaps (over 30 days) is low (1,8%). In total, customers spent 1.220,680.44 monetary units the 3 months, with the highest individual spend being 1418.33. Due to outliers, we applied a log transformation to the total_spent variable, resulting in a normalized distribution. The average amount spent is approximately 24.25 monetary values, with a standard deviation of 2.63, and the value for the third quartile is 45.18. This suggests that despite some high-spenders, most customers spend much less. For "order_frequency", the mean is approximately 1, showing an average of 1 order per week. The maximum number is approximately 28 orders per week. This feature is useful to identify high-frequency customers that can be targeted with loyalty programs, special promotions, or personalized offers to maintain their engagement and encourage continued spending. We also analyzed the number of unique cuisines per customer to find that around 72% of customers stick to 1 or 2 types of cuisine, with an average of 2 unique cuisines per customer and a maximum of 13.

Regarding new categorical features we can see their distribution in figure 8 and summary in table 6. For customer activity, most customers (51.1%) are active, while 34.48% are churning, and 14.42% are recent. Even though most customers are active, the significant number of churning customers shows potential areas for improving retention. For "promo_used", as stated earlier, there is a balance between the number of customers using a promo in the last order (around 47%) and those who have not (53%). For "favorite_cuisine" (based on spending) we confirmed earlier observations. We created a new feature "favourite_cuisine" to identify each customer's top cuisine based on their spending. Confirming earlier observations, Asian cuisine stands out as top choice, with 22% (7021) of customers spending the most on it. Interestingly, while Noodle Dishes contribute the least to total spending, they rank 11th, and Cafe ranks as the least favourite, chosen only by 1.5% as their top cuisine. For "favorite_day" (the day that each customer placed more orders), Sunday is the mode, with 20% of customers placing orders on this day, making it the most popular day, even though it does not have the highest number of total orders. This means that although most customers prefer Sunday, they usually place fewer orders in total that day (making the sum of it lower). For "favorite_hour" (the hour each customer placed more orders), we see a similar distribution with 3 peaks (11H; 16H; and 3H).

3.2. Missing values and Outliers

We found **missing values** in "customer_age", "first_order", "HR_0" and "customer_region" ("-"). For "first_order", we verified that when missing, most cases had only 1 order (only 2 cases with 2 orders), so we concluded that that first order day would be the same as the last order day, like it happened in other instances with only 1 order. For the 2 cases with 2 orders, we assumed that they were on the same day. For "HR_0" imputation, we calculated the difference between total orders (sum of all orders for each day) and the total orders from HR-1 to HR-23. For “customer_age” we used KNN imputation (after scaling the data with Robust Scaler which is more resistant to the outliers present in our data). For “customer_region” imputation, we converted missing values described as “-” to np.nan, then converted the categories to integers with Label Encoder to use KNN Imputer on the new numerical variables (also scaled with Robust Sclaer). These strategies are summarized in table 7.

For outliers, since extreme outliers don’t reflect the central tendency, we removed them (especially when the variable was extremely skewed due to them). In table 8 in the annexes, we show each cutoff for each selected feature. After removing outliers, we retained 97,3% of the original dataset size, removing 857 customers (rows).

4. MULTIVARIATE ANALYSIS

For **comparison between two numerical features**, we built a correlation matrix and pair plots to identify associations. We used Spearman's correlation coefficient as it is more resistant to outliers and better identifies non-linear relationships than Pearson's correlation (figure 9). We found a strong positive correlation between "total_orders" and "vendor_count" (0.89), and also with "product_count" (0.95), so for feature analysis, we dropped the last two variables. Since “percentage_chain” and “is_chain” represent the same concept, we dropped “is_chain”. Due to many features, refer to table 9 in the annexes for a summary description of positive and negative correlations. Key findings: higher spending correlates with buying at unusual hours (3-5; 8-11); More expensive cuisines: Asian; Street Food; Cheaper ones: Chicken; Noodles; Other. Spending on Asian cuisine negatively correlates with spending on Italian cuisine (figure 10). Buying at hour “x” correlates with next hour, and buying at lunch correlates with dinner time (figure 12). Frequent app users place lower average-value orders; lower average value per correlates with a higher percentage of chain spending (suggesting most users who spends less money, buy chain food). Italian Cuisine has fewer chain companies. A positive correlation between “total_orders” and “average_time_between_orders” is observed in the pairplot (figure 13) and density pairplot (figure 14), showing that the positive correlation, meaning weekly orderers place more total orders. In the density pair plot between “customer_age” and “total_spent” (log), younger ages (< 7 CUI) and (> 54 CUI) spend less, while ages > 40 spend median amounts of money (7 - 54 CUI in 3 months).

For the **comparison of two categorical features**, we built bar plots with relative proportions for the different categories. See table 10 in the annexes for summarized insights. Comparing "favorite_cuisine" with "customer_region", regions starting with the same number preferred the same food (8: Asian; 4: Italian; 2: Other), so we grouped them by their first digit (figure 15). Recent and churning customers often have recent promotions (mainly delivery promotions) in the last order, probably to retain them (figure 17).

For **comparison of a categorical to a numerical feature**, we built box plots. See table 11 in the annexes for a summary. Region 2 mainly spends on chain, buying cheaper orders, spending less, mostly on Indian and Other Food (figure 18). Region 8 and 4 spend less on chain food but more on the app, with higher average order prices and different preferences (8: more on Street Food and Asian; 4: more Chinese and Indian) (figure 19 and figure 20). Delivery discounts are mainly applied to recent customers with low “lifetime”, while freebies are applied to older customers (higher lifetime) (figure 21). High spenders on Italian food buy less from chain companies; desert buyers have shorter lifetime (possibly occasional buyers). Customers preferring Street Food and Asian spend more money overall in the app (also with higher average price per order) and those who prefer Chicken and Noodles spend less money on the app (and also lower average price per order) (figure 22).

For **comparison between two numerical features and one categorical feature**, we built density plots between “customer_age” and (log) money spent per cuisine, between different regions (figure 23.1 and figure 23.2). See table 12 in the annexes for summary and insights. For region 8 there are usually 2 customer centres between 20-30 years: higher spenders, maybe richer, and lower spenders. Other regions show more even distributions. There is also some segmentation across age for some regions: there are clusters of customers in region 8, between 40-50 years that spend a relatively high amount of money on Chinese cuisine. In region 4 there are 2 clusters of customers: one with customers between 30-40 years that spend a high amount of money on Healthy Food; and the other with customers between 50-60 years that spend the median amount of money in Indian and Thai food. For the “2” region there is one cluster of customers of 50-60 years that spend a median amount of money on Beverages.

For **co-occurrences analysis**, we created co-occurrence tables for cuisine types (table 13), days of the week (table 14), and hours of the day (table 15). Each co-occurrence table shows how often two variables appear together. Most cuisine types are bought together with "American", "Asian" and "OTHER" (figure 24). For hours of the day, most of the orders are placed within the next hour.

5. DISCUSSION AND CONCLUSIONS

Based solely on the exploratory analysis, we can start to cluster the customers according to their value (monetary spent), demographics, and preferences, and suggest some marketing strategies regarding each cluster. Examples would be following co-occurrence frequency for recommendation systems; or marketing for specific clusters that were presented (age groups; regions), or based on preferences (day preference; hour preference). Future steps will involve applying clustering algorithms to segment customers, enabling ABCDEats to optimize its marketing efforts and drive business growth.

BIBLIOGRAPHICAL REFERENCES

Dolnicar, S., Grün, B., & Leisch, F. (2018). *Market segmentation analysis: Understanding it, doing it, and making it useful*. Singapore: Springer Nature. https://doi.org/10.1007/978-981-10-8818-6

OpenAI. (2024). ChatGPT. Available at https://openai.com. ChatGPT was utilized to assist in developing certain sections of the code (figures coding mainly).

ANNEXES

Table 1 - Overview of Dataset:

Data Overview	
Number of variables	56
Number of observations	31888
Missing cells	1998
Missing cells (%)	6.2 %
Duplicate rows	13
Duplicate rows (%)	0.04%
Variables types	
Numeric	52
Categorical	4

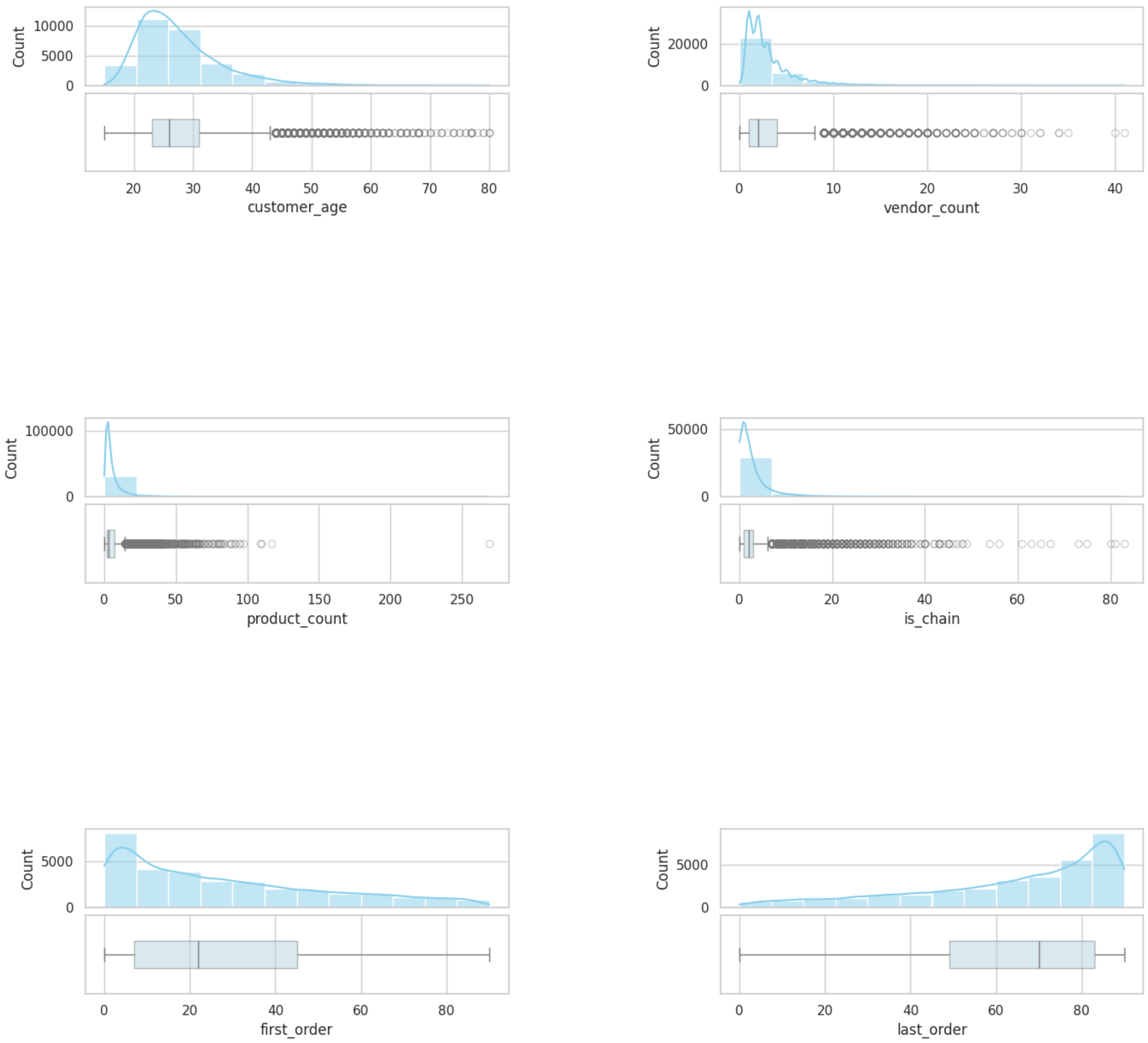

Figure 1 - Initial numerical variables histograms and boxplots

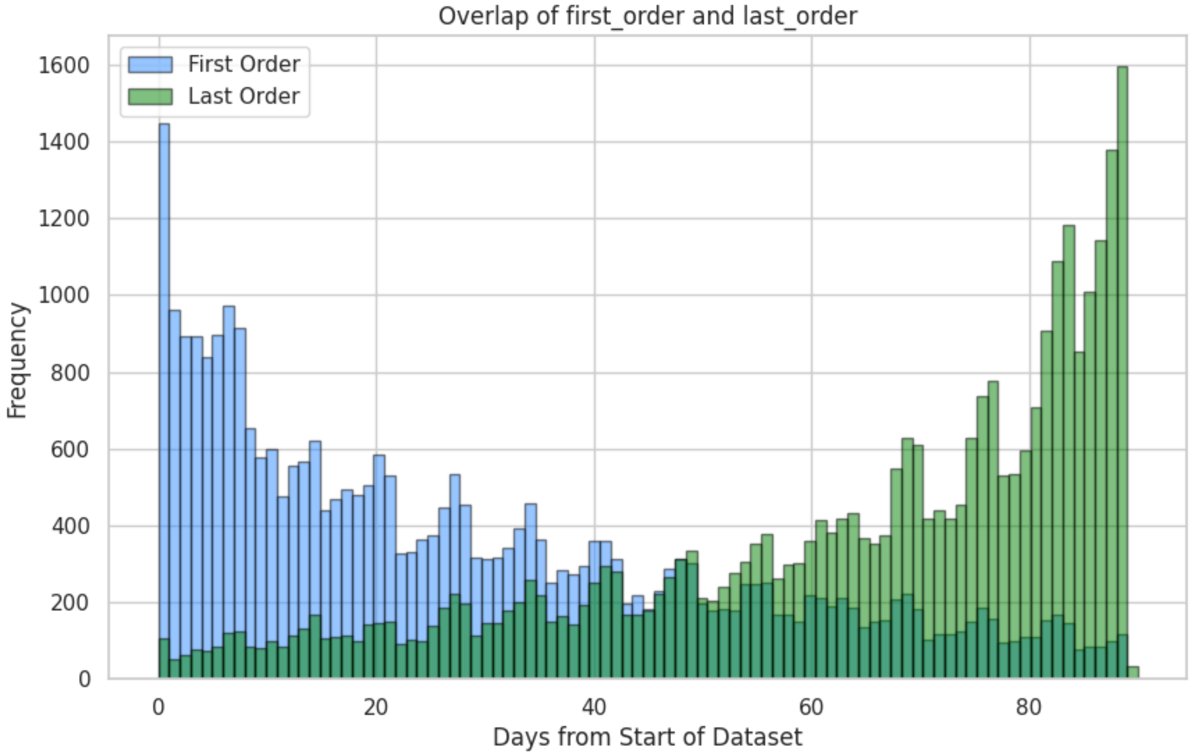

Figure 2 – Distribution of First and Last order

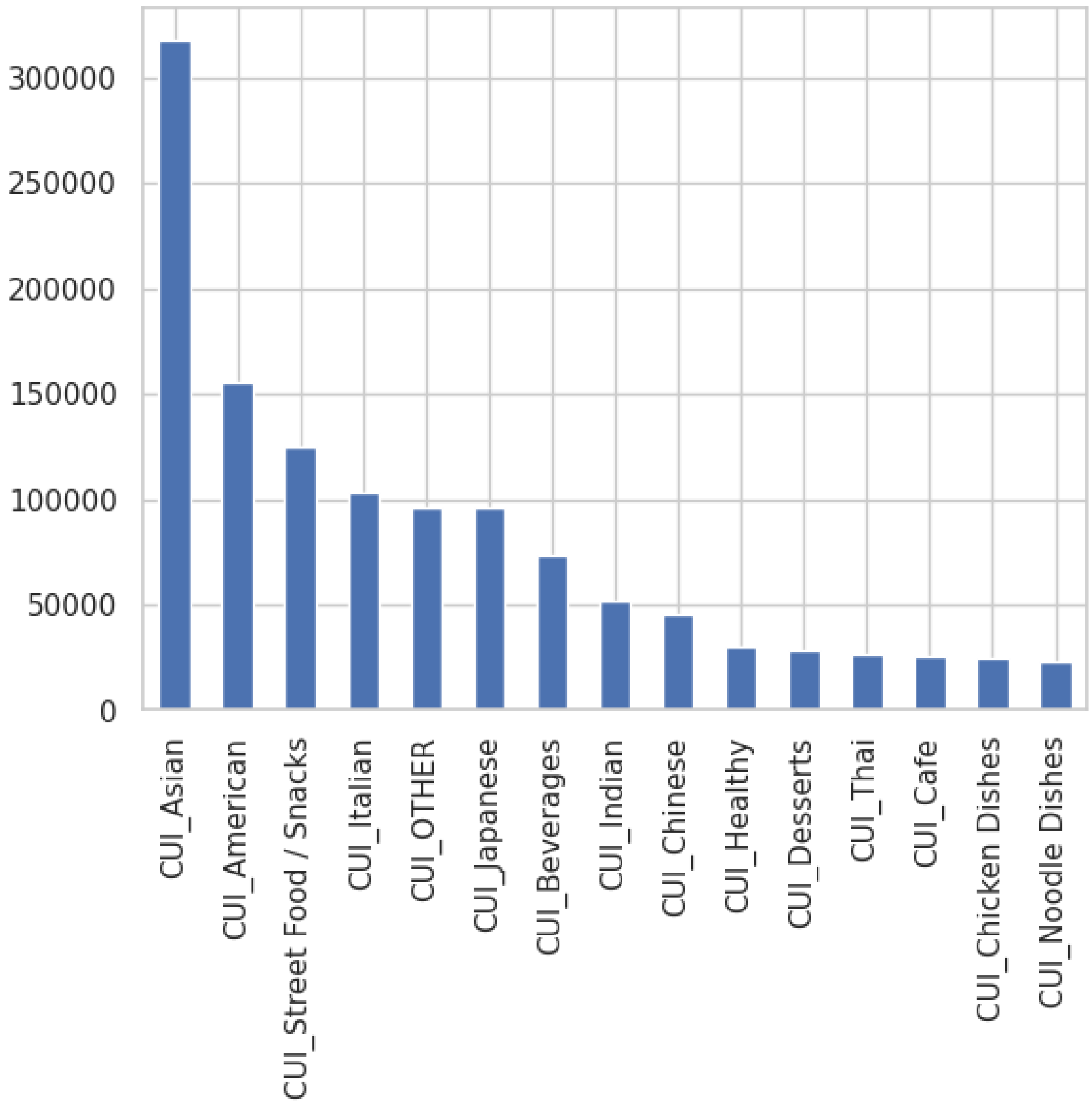

Figure 3 – Distribution of sum of CUI per cuisine

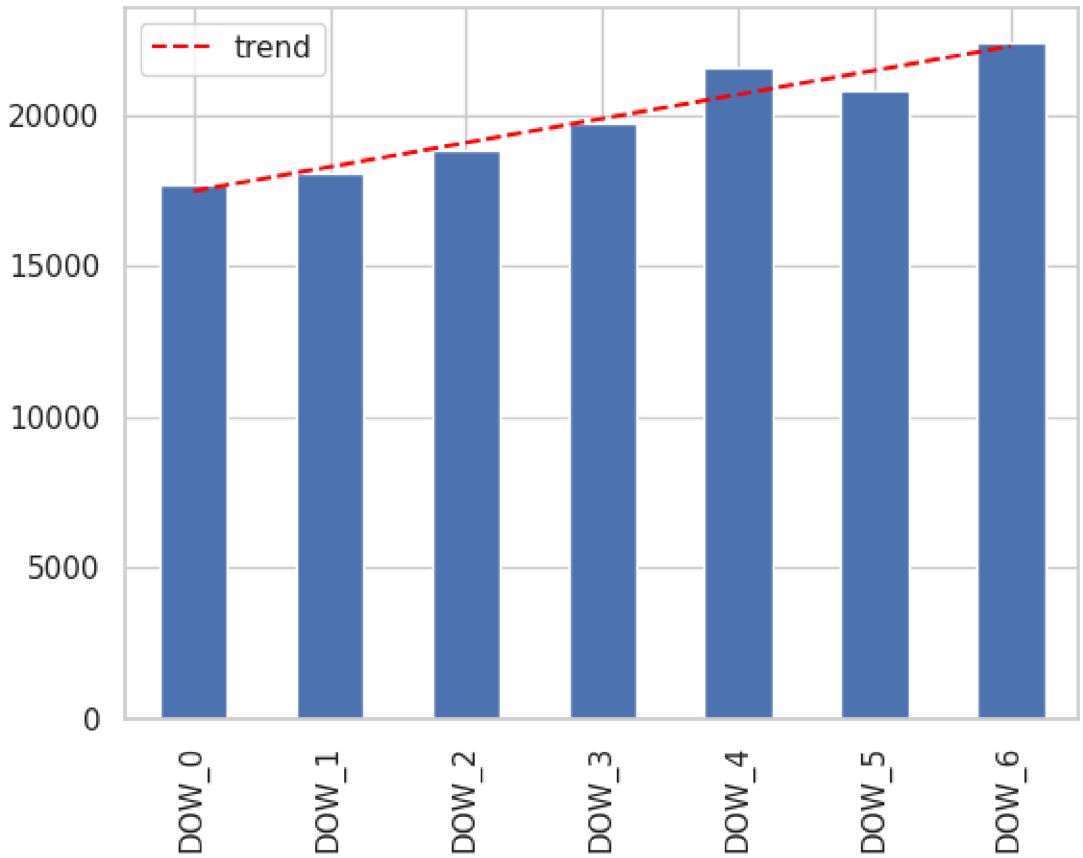

Figure 4 - Distribution of sum of total orders per days

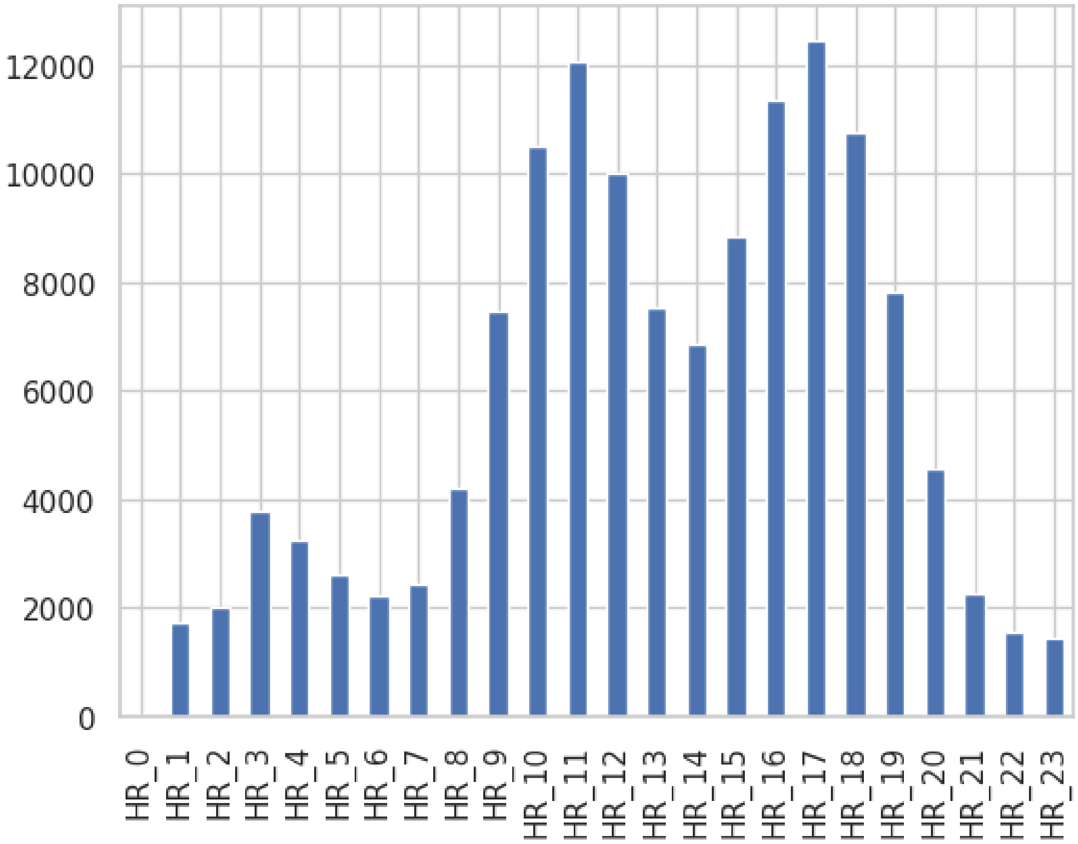

Figure 5 - Distribution of sum of total orders per hours

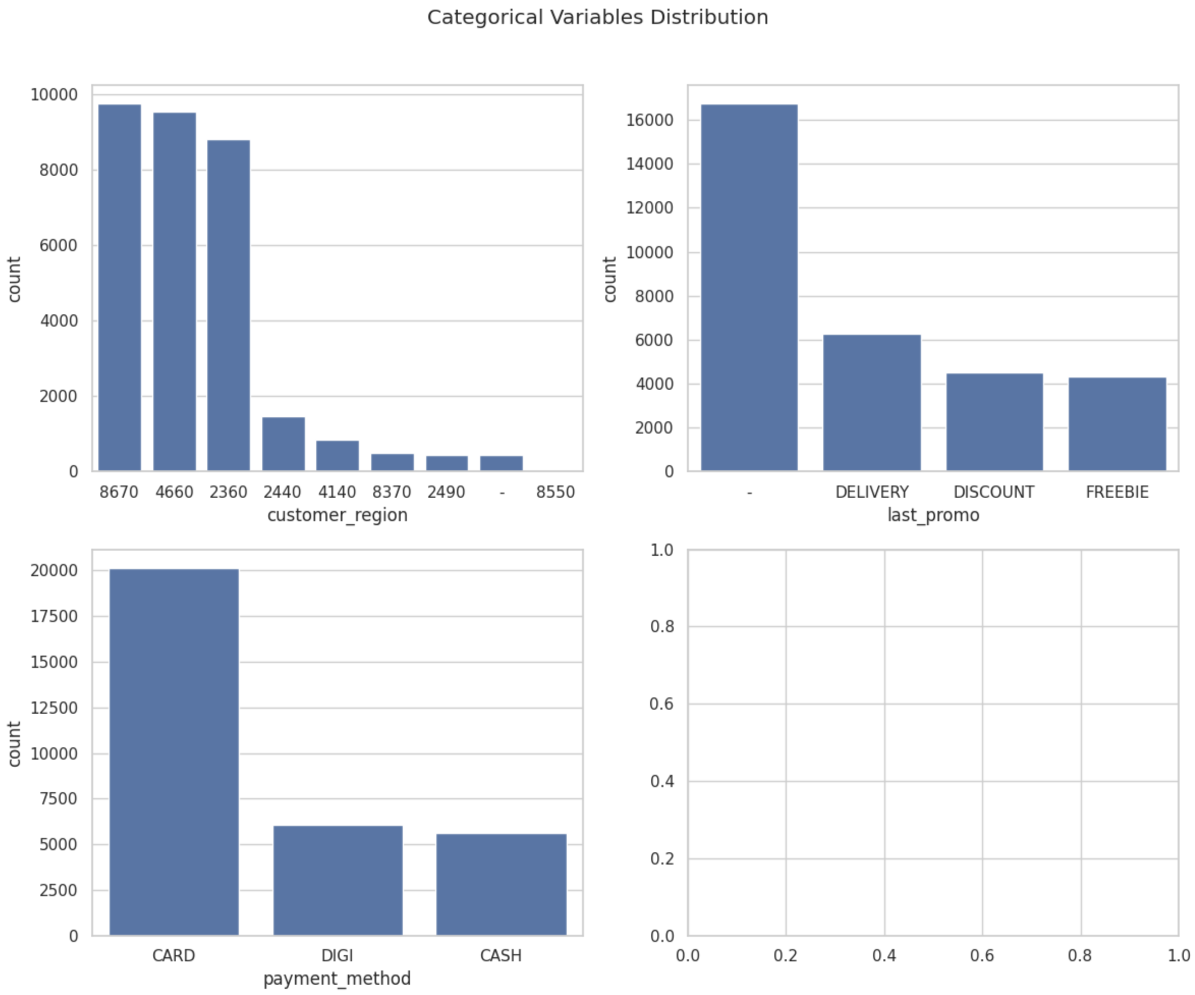

Figure 6 - Initial categorical variables distribution – bar plots

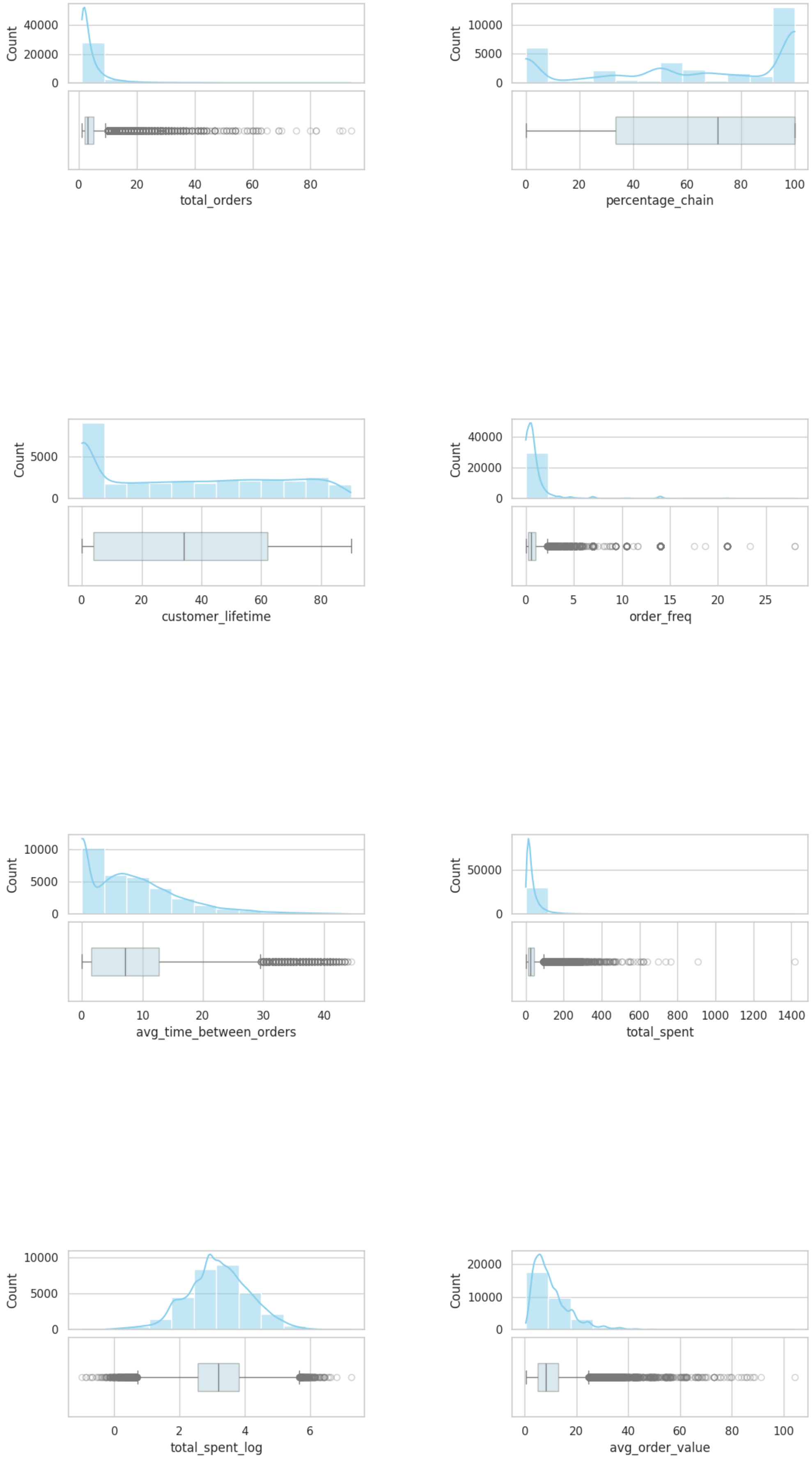

Figure 7 - New numerical features distribution histograms and boxplots

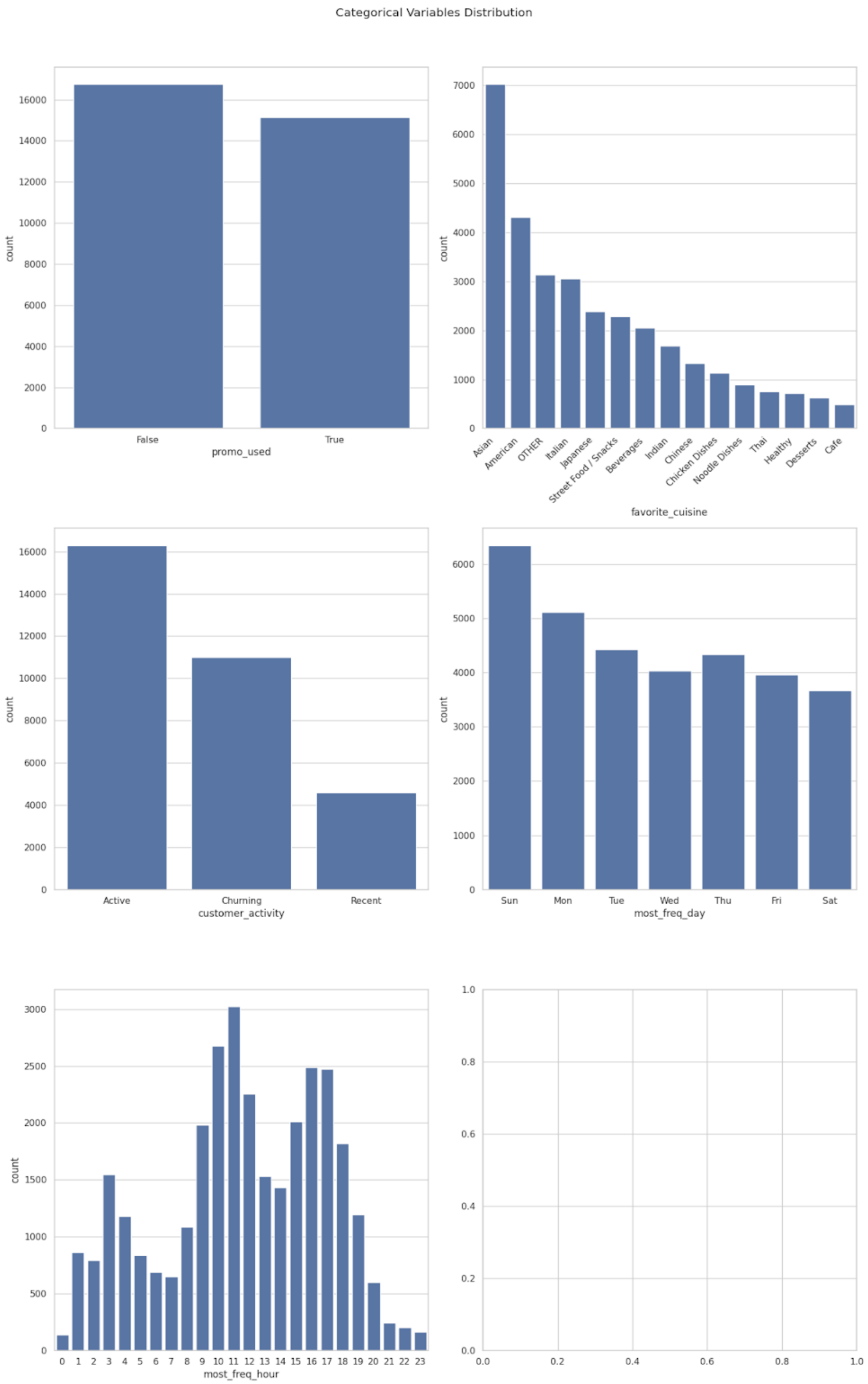

Figure 8 - New categorical features variables bar plots

Table 2 - Statistical summary of initial numerical variables

Feature	Mean (+- std)	median	Q25-Q75	Min-max
Customer_age	28 ± 7	26	23 - 31	15 - 80
Vendor_count	3 ± 3	2	1 - 4	0 - 41
Product_count	6 ± 7	3	2 - 7	0 - 269
Is_chain	3 ± 4	2	1 - 3	0 - 83
First_order	29 ± 24	22	7 - 45	0 - 90
Last_order	64 ± 23	70	49 - 83	0 - 90
CUI_<type>				
DOW_<0-6>	4 ± 5	3	0 - 5	0 - 94
HR_<0-23>	4 ± 4	3	0 - 5	0 - 39

Table 3 - Distribution of categorical variables by class

Feature	Distribution
Customer_region	“8670” (30,6%) “4660” (30,0%) “2360” (28,0%) “2440” “4140” “8370” “2490” “8550” “-”
Last_promo	“-” (53%) “DELIVERY” (20%) “DISCOUNT” (14%) “FREEBIE” (14%)
Payment_method	CARD (63%) DIGI (19%) CASH (18%)

Table 4 - Description of new features and how they were built

New feature	Feature description	Data type	How it was created
Total orders	Total orders made by each client	Numerical (int)	The sum of orders for each day.
Percentage_chain	Percentage of orders that were from Chain companies	Numerical (int)	Division of "is_chain" orders for "total_orders", multiplied by 100
Customer activity	Classification for customer activity	Categorical (string)	"Recent": first_order >= days "Active": first_order < 60 days & last_order >= 60 days "Churning": first_order < 60 days & last_order < 60 days
Customer_lifetime	Number of days between the last and first order	Numerical (int)	Subtraction between "last_order" and "first_order"
Order frequency	Number of orders per week during customer's lifetime	Numerical (int)	The total number of orders divided by customer lifetime, then divided by seven (days).
Avg_time_ between_orders	Number of days between each order during the customer's lifetime (assuming the client ordered with a fixed time interval)	Numerical (int)	The customer's lifetime is divided by the total number of orders.
Promo_used	If promo was used or not	Boolean	Agrupating "last_promo" to: False if ("-"); True, if ("DELIVERY", "DISCOUNT", "FREEBIE")
Total_spent	Total money spent in the app during the follow-up	Numerical (float)	Summing the number of monetary units spent by the customer for every type of cuisine: "CUI_<cuisine_type>"
Avg_order_value	Average monetary units spent in each order	Numerical (float)	Total spent divided by the total number of orders
Favorite_cuisine	Cuisine associated with the highest amount spent by the customer.	Categorical (string)	Assign the cuisine type associated with the highest amount spent for each respective customer

Favorite_day	Day of the week with the highest number of orders placed by the customer	Categorical (string)	Assign the day number associated with the highest number of orders for each respective customer
Favorite_hour	Hour of the day with the highest number of orders placed by the customer	Categorical (string)	Assign the hour number associated with the highest number of orders for each respective customer

Table 5 - Statistical summary of new numerical variables

Feature	Mean (+ std)	Median	Q25-Q75	Min-max
Total_orders	4 ± 5	3.0	2 - 5	0 - 94
Percentage_chain	62 ± 39	71	33 - 100	
Customer_lifetime	35 ± 29	34	3 - 62	0 - 90
Order_frequency (weekly)	1 ± 2	1	0 - 1	0 - 28
Avg_time_between_orders	8 ± 8	7	1 - 12	0 - 44
Total_spent	38,2 ± 46,4	24,1	12,8 - 44,9	0 - 1.418,3
Avg_order_value	10,2 ± 7,8	8,1	5,0 - 12,9	0 - 104,3

Table 6 – Distribution of new categorical features

Feature	Distribution
Customer_activity	“Active” (51%) “Churning” (34%) “Recent” (14%)
Promo_used	False (53%) True (47%)
Favorite_cuisine	Asian (22%) American (13%) Other (10%) Italian (10%) Japanese (8%) Street Food / Snacks (7%) Beverages (6%) Indian (5%) Chinese (4%) Chicken (4%) Noodle (3%) Thai (2%) Healthy (2%) Deserts (2%) Café (2%)
Favorite_day	Sunday (20%) Monday (16%) Tuesday (14%) Wednesday (13%) Thursday (14%) Friday (12%) Saturday (12%)
Favorite_hour	HR_0 2% HR_1 2% HR_2 2% HR_3 5% HR_4 4% HR_5 3% HR_6 2% HR_7 2% HR_8 3% HR_9 6% HR_10 8% HR_11 9% HR_12 7% HR_13 5% HR_14 4% HR_15 6% HR_16 8%

	HR_17 8% HR_18 6% HR_19 4% HR_20 2% HR_21 1% HR_22 1% HR_23 0%

Table 7 - Dataset missing values imputation strategy

Feature	Missing values	Imputation strategy	Justification / Steps
“first_order”	106	Impute corresponding value of “last_order”	Verify that the number of total orders where first_order is missing ++ 1 order, meaning the first and last order days are the same.
“HR_0”	1165	Impute the difference between total orders and total orders excluding “HR_0”	Sum of total orders in every week (total orders during analysis). Sum of total orders for HR_1 - HR_23 Missing value of HR_0 are the difference between total orders and total orders HR_1-HR_23
“customer_age”	727	KNN Imputer (5 neighboors)	Robust Scaling of numerical features (more resistant to outliers) KNN Imputation
“customer_region”	442	KNN Imputer (5 neighboors)	Transform “-” to np.nan Label encode categories in this feature (transforming to numerical) Robust Scaling KNN imputation to missing (np.nan)

Table 8 – Outliers cutoff for numerical features

Features	Cutoff
Customer age	55
Vendor count	50
Is chain	30
Total orders	40
Order freq	8
Avg time between orders	40
Avg order value	70
Total spent	300

Table 9 - Numerical variables correlation

Numeric variable		Positive correlation	Negative correlation
Total orders		Customer lifetime Order frequency Hours 8-20 Avg time between order *Cuisines*: American / Asian / Chicken / Chinese / Italian / Japanese / Other	Avg order value
Total spent		Avg order value Cuisines: American / Asian / Japanese / Street Food Hour: 3-5 / 8 - 11/ 14-16	Percentage chain
Order frequency		*Cuisines*: American / Other	
Average order value		Total spent *Cuisines*: Asian / Street Food	Total orders Percentage chain Hour 18-20 *Cuisines*: Chicken, Noodle, Other
Percentage chain			Avg order value Total spent *Cuisines*: Italian
Cuisines	Asian		Italian
	Chicken	Chinese, Noodles, Other	
Days of week	Sunday	Monday - Thursday	
Hours	Next hour	Previous hour	
	12-13	18-19	

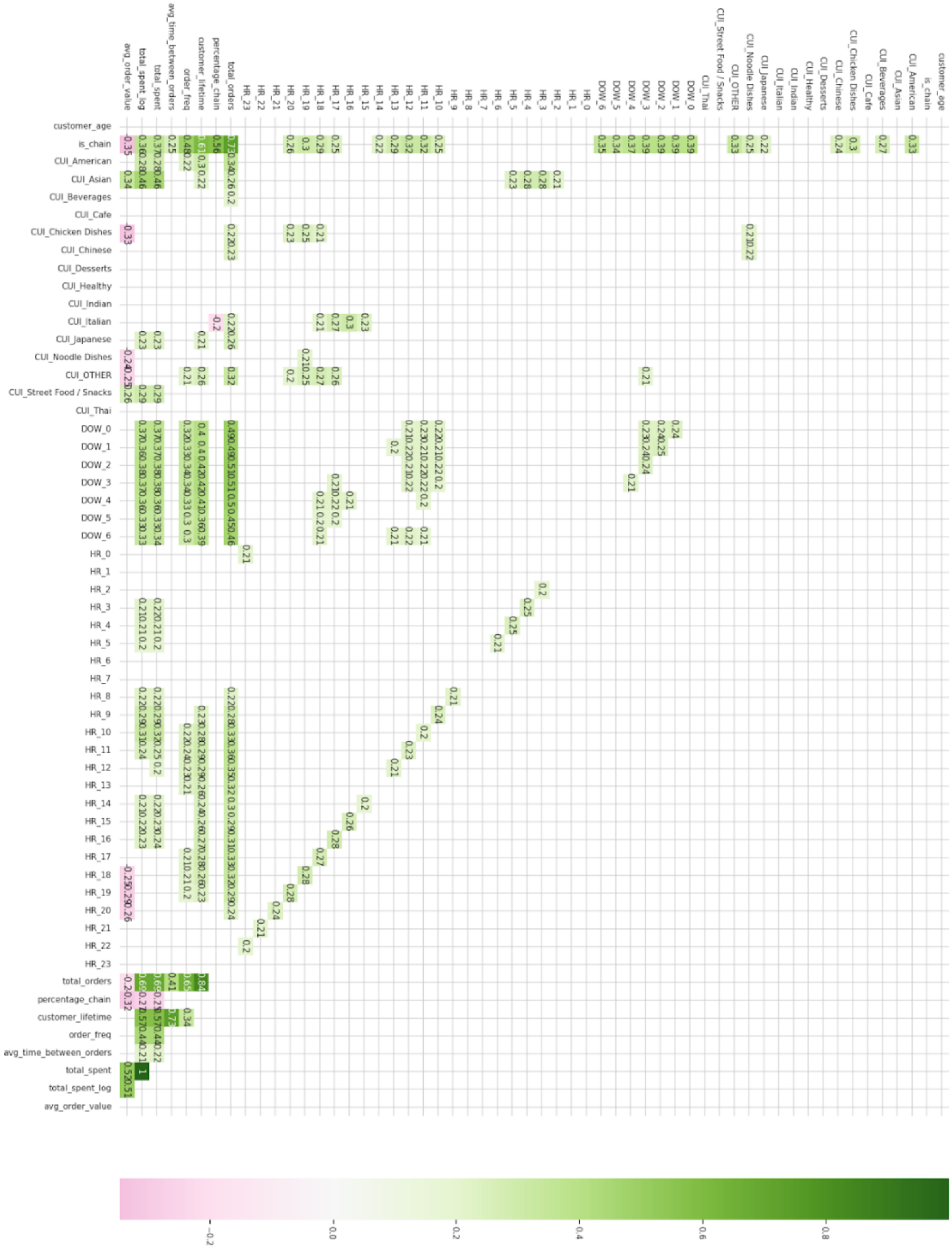

Figure 9 - Correlation matrix (numerical variables)

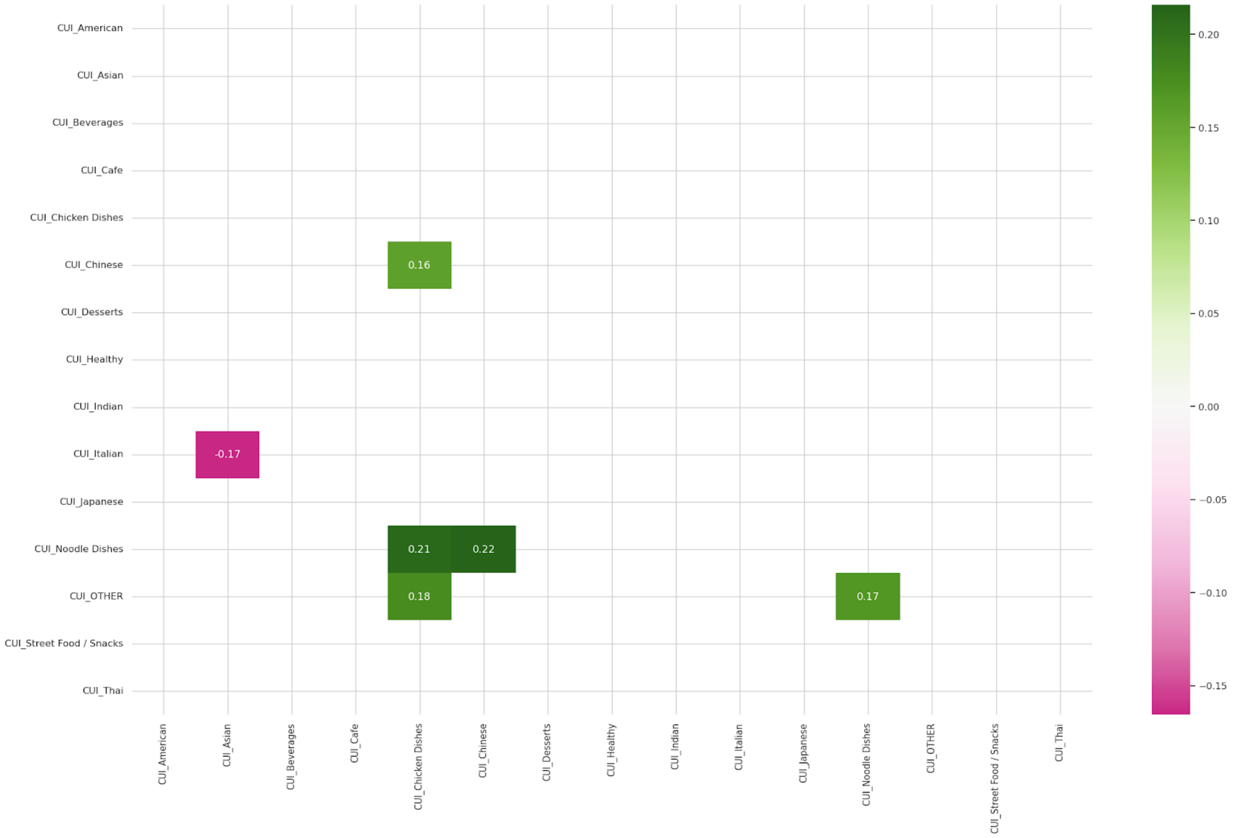

Figure 10 - Cuisines correlation matrix

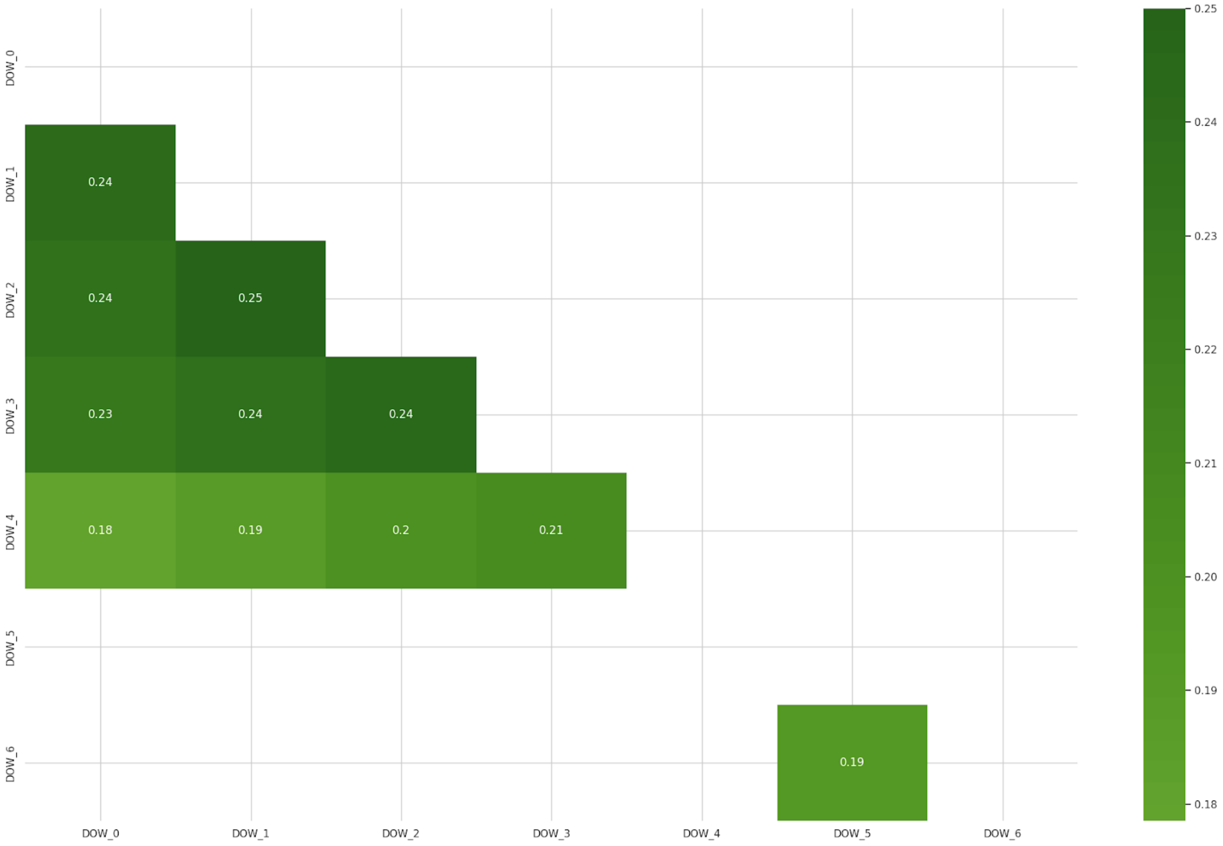

Figure 11 - Weekdays correlation matrix

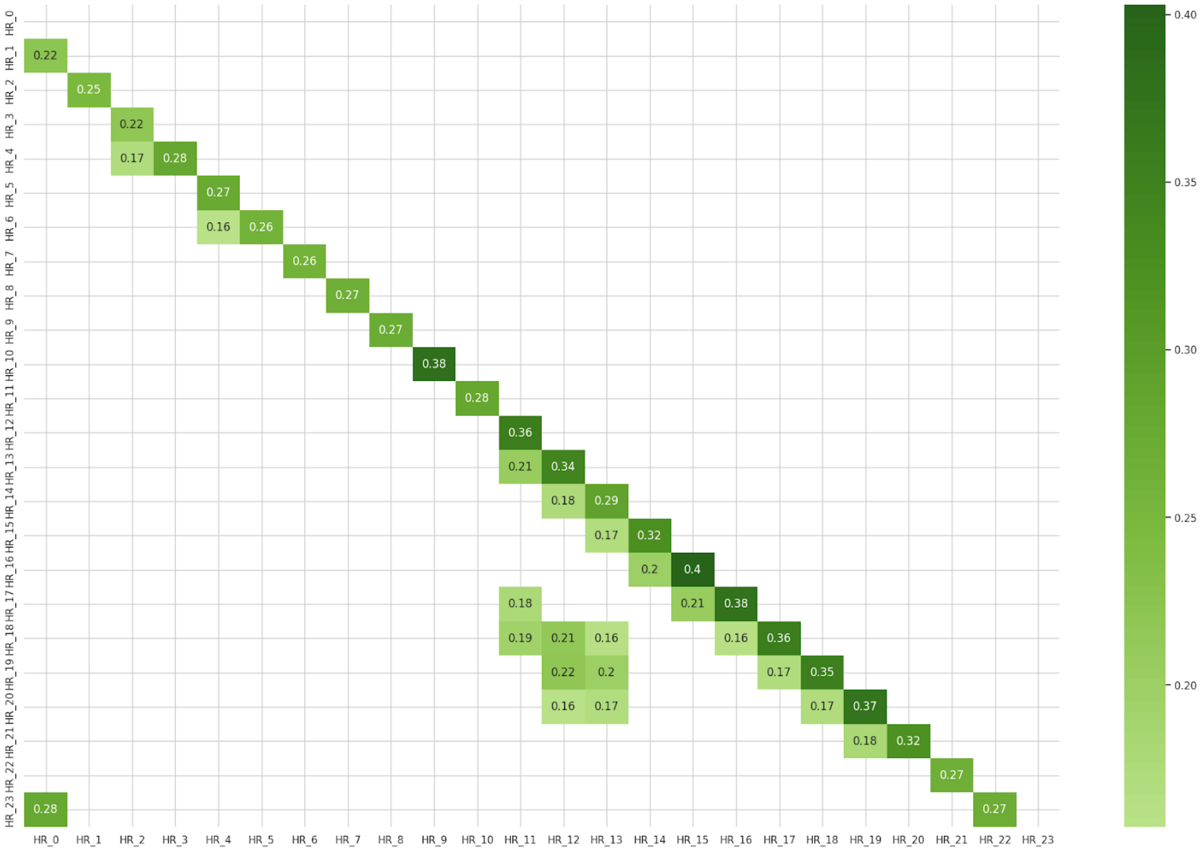

Figure 12 - Hours correlation matrix

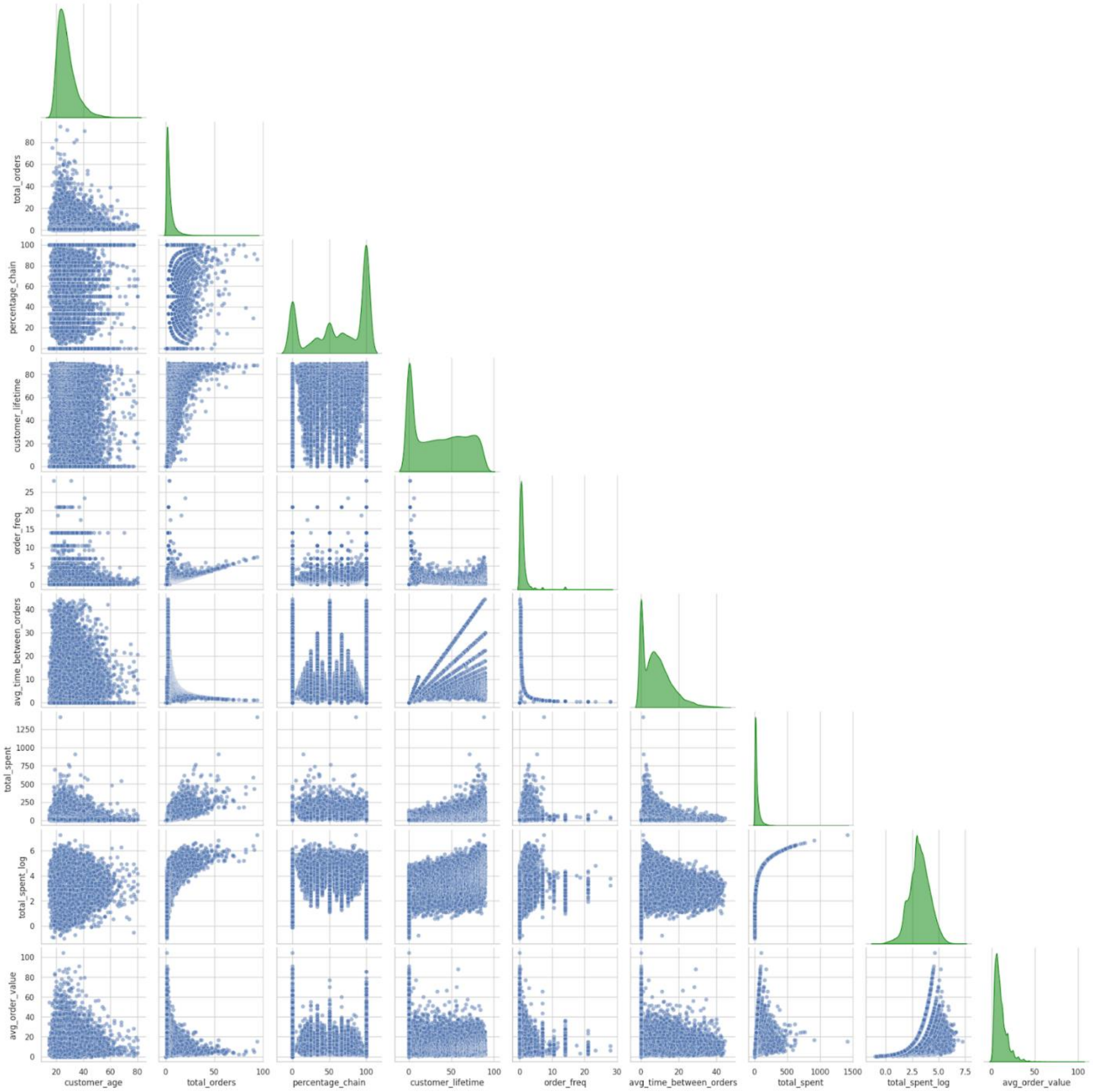

Figure 13 – Numerical features pairplots

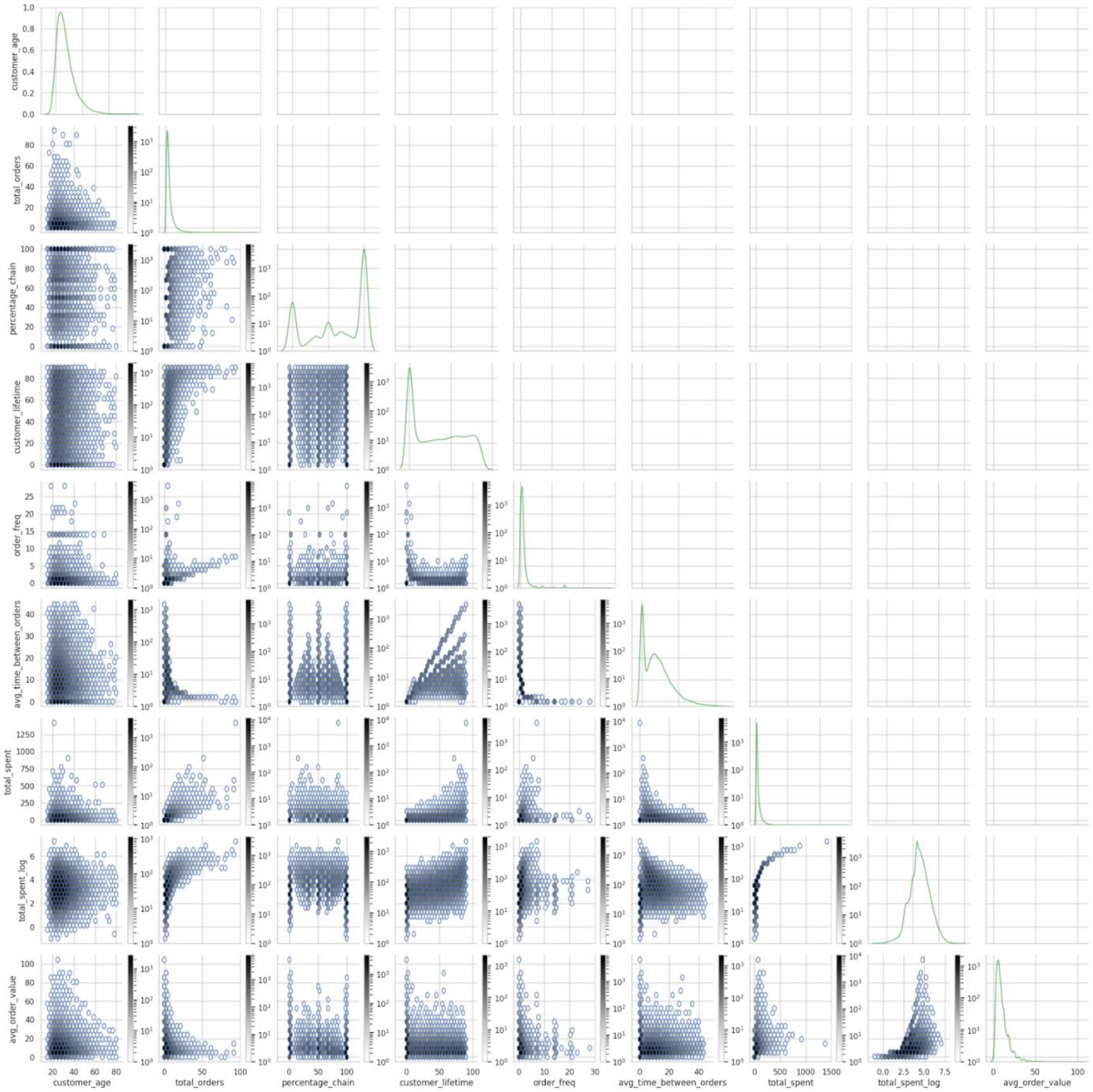

Figura 14 – Numerical features density pairplots

Table 10 – Categorical features comparisons

1st categorical variable		2nd Categorical variable	
Customer Region	Starting with 8	Asian	Favorite food (more money spent on)
	Starting with 4	Italian	
	Starting with 2	Other	
Most frequent hour (most orders placed in)	1-4	Asian ++	Favorite food (more money spent on)
	6-7	Beverages (some)	
	7-9	Cafe (some)	
Customer activity	Recent	Promo used ++ (delivery ++)	Last promo
	Churning	Promo used + (delivery ++)	
	Active	No promo (++)	

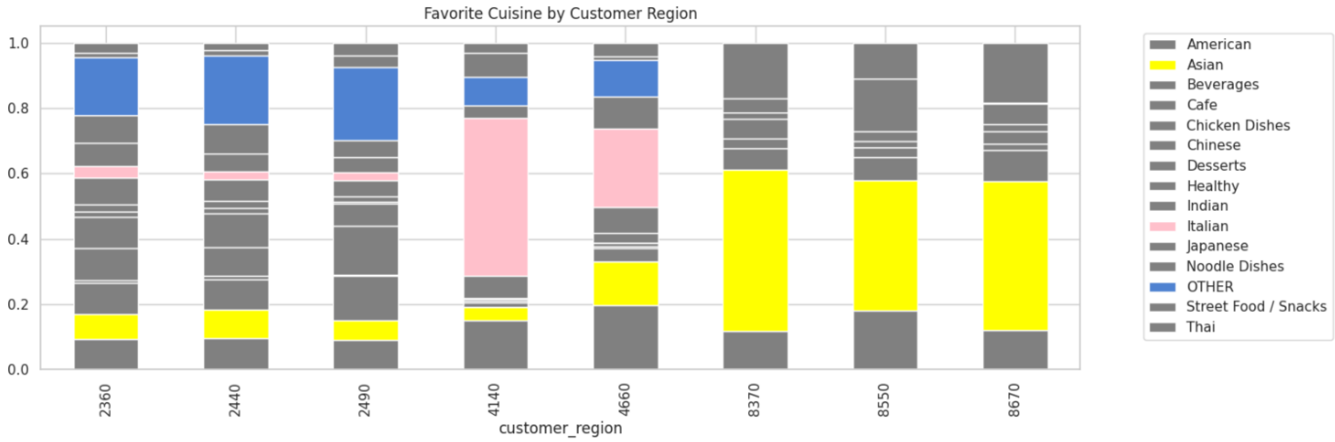

Figura 15 - Favorite cuisine distribution per customer region.

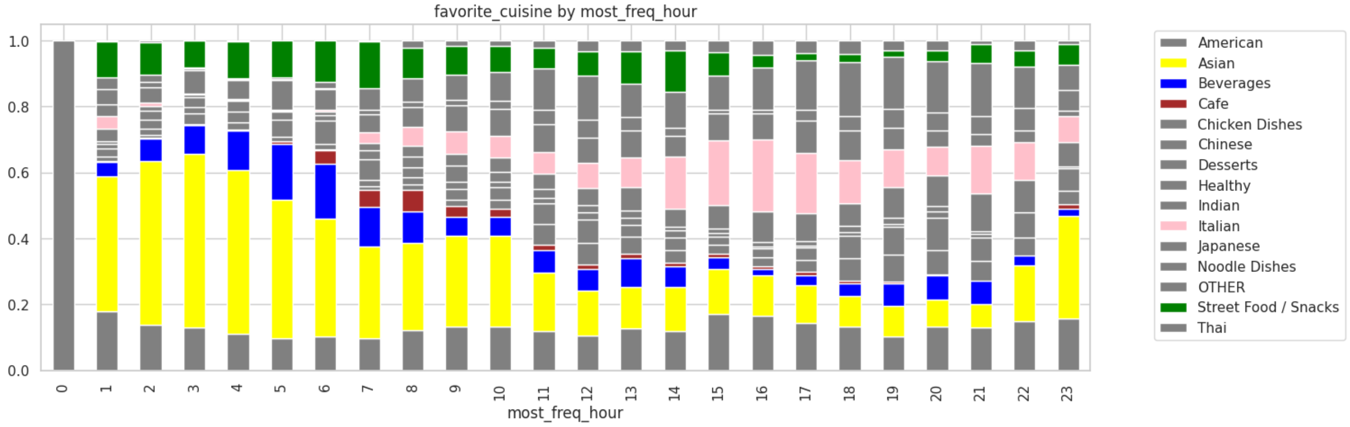

Figure 16 - Favorite cuisine distribution per favorite hour.

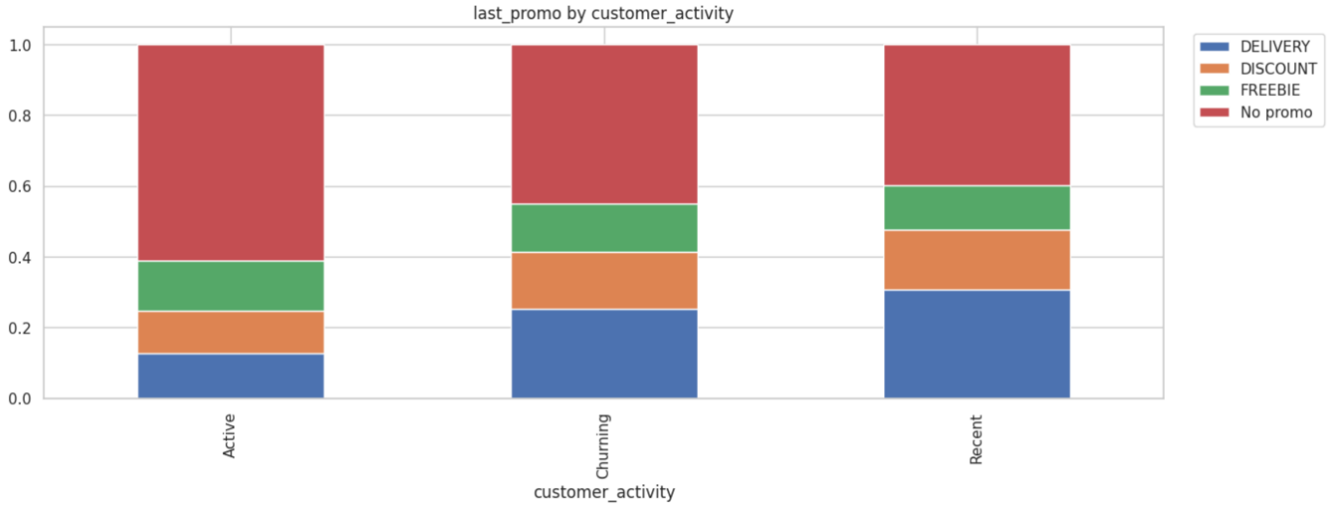

Figure 17 - Last promo distribution per customer activity.

Table 11 - Categorical and numerical variables assocations

Categorical	Numerical	Association * (the numerical values here presented represent the median)
Customer region	Percentage chain	2: 100% chain 8 and 4: 50% chain
	Total spent (log)	8: 3,6 (37 CUI) 4: 3,3 (27 CUI) 2: 2,7(15 CUI)
	Average order value	8: 13 CUI 4: 9 CUI 2: 5 CUI
	Money spent (log) each cuisine	Median 8: Street Food (3,2= 25 CUI)/ Asian (3,1= 22 CUI) 4: Chinese (2,7 = 15 CUI) / Indian (2,6 = 13 CUI) 2: Indian (1,9 = 7 CUI) / Other (1,8 = 6 CUI)
		Total: 8: Asian (44%), Street Food (20%) 4: Italian (23%), American (18%) 2: Other (17%), Chinese (10%)
Last promo	Customer lifetime	No promo: 41 days Freebie: 38 days Discount: 27 days Delivery: 7 days
Favorite cuisine	Percentage chain	Noodle / Chicken / Beverages / American: 100% Italian: 19%
	Customer lifetime	Others: 30-40 days Desert: 20 days
	Average order value	Street Food: 15 CUI Asian / Deserts / Japanese: 10 CUI Noodle / Chicken: 5 CUI
	Total spent (log)	Street Food: 3,6 (37 CUI) Asian / Cafe: 3,4 (30 CUI) Noodle: 2,9 (18 CUI) Chicken: 2,5 (12 CUI)

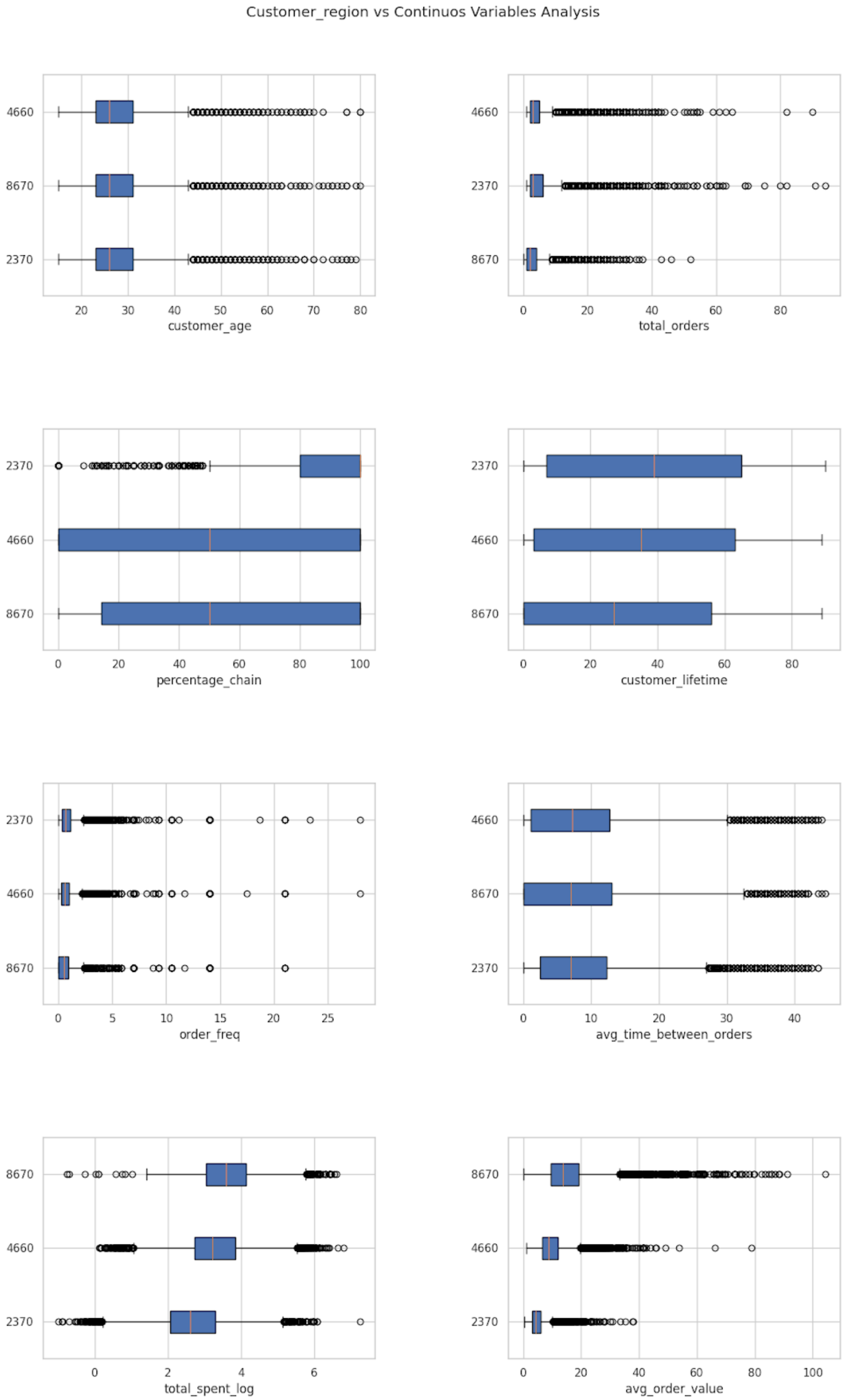

Figure 18 - Customer region vs continuous variables analysis

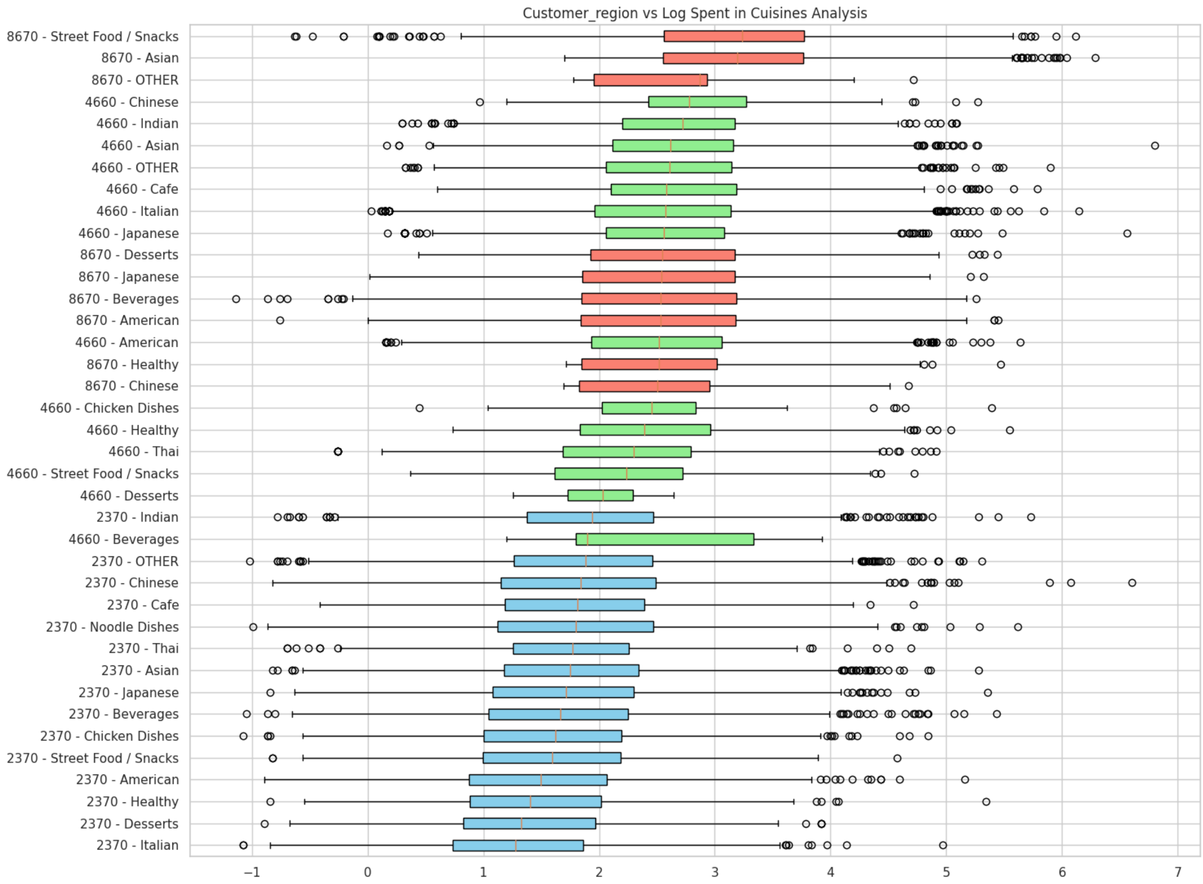

Figure 19 - Customer region vs. natural log spent in each cuisine.

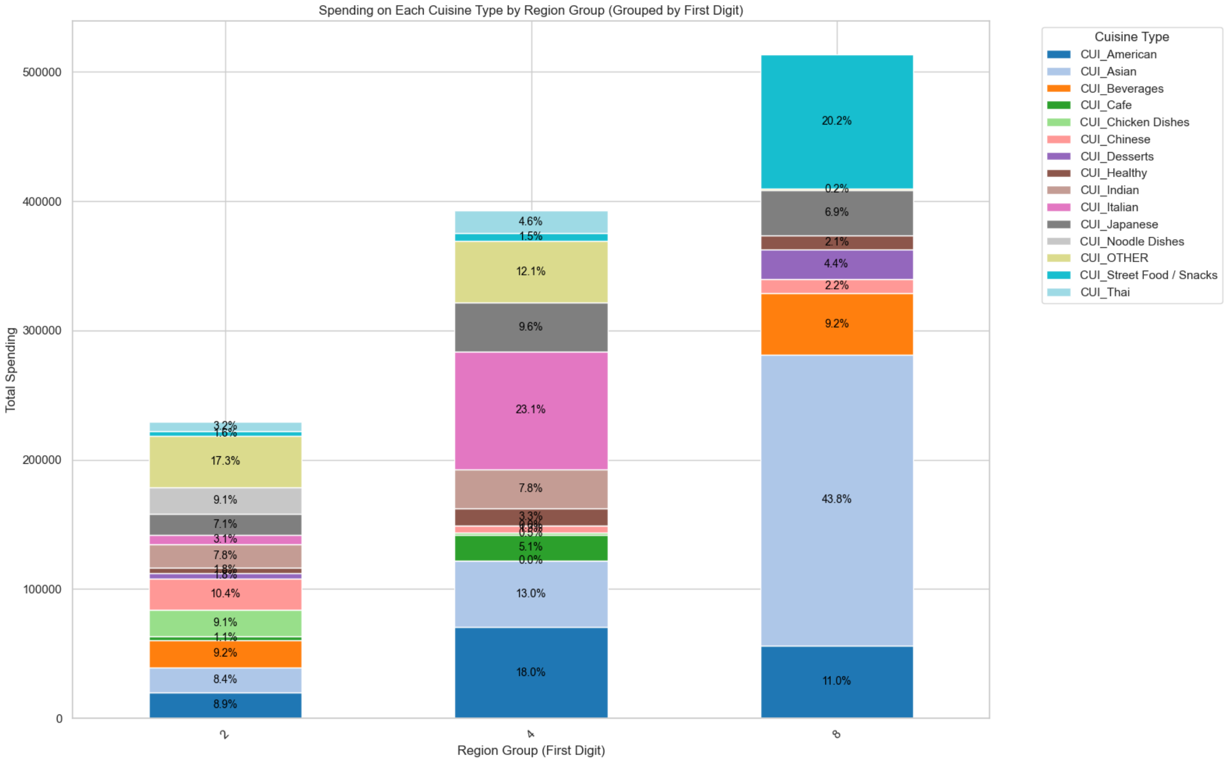

Figure 20 - Spending in each cuisine type by region group.

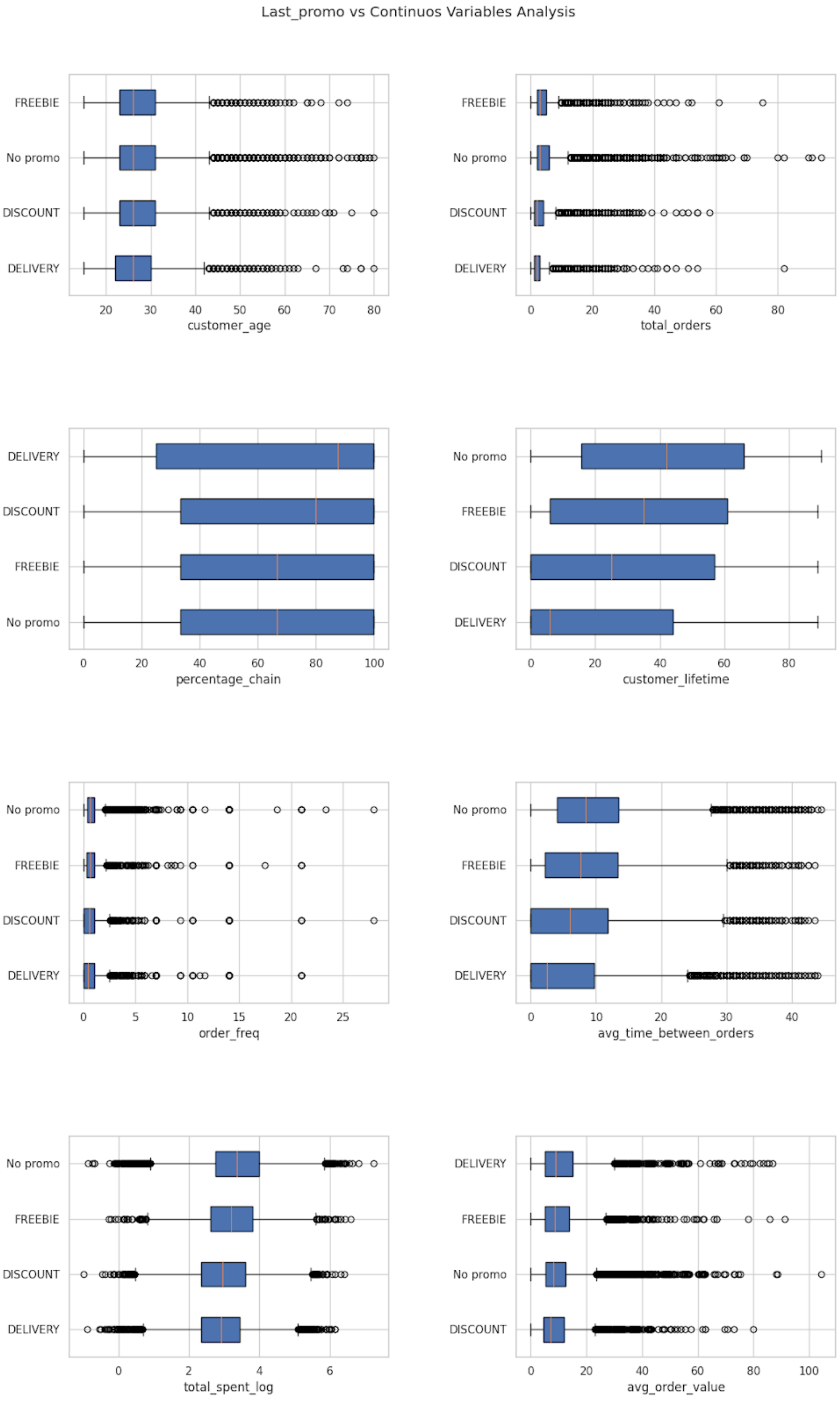

Figure 21 - Last promo vs numerical variables analysis

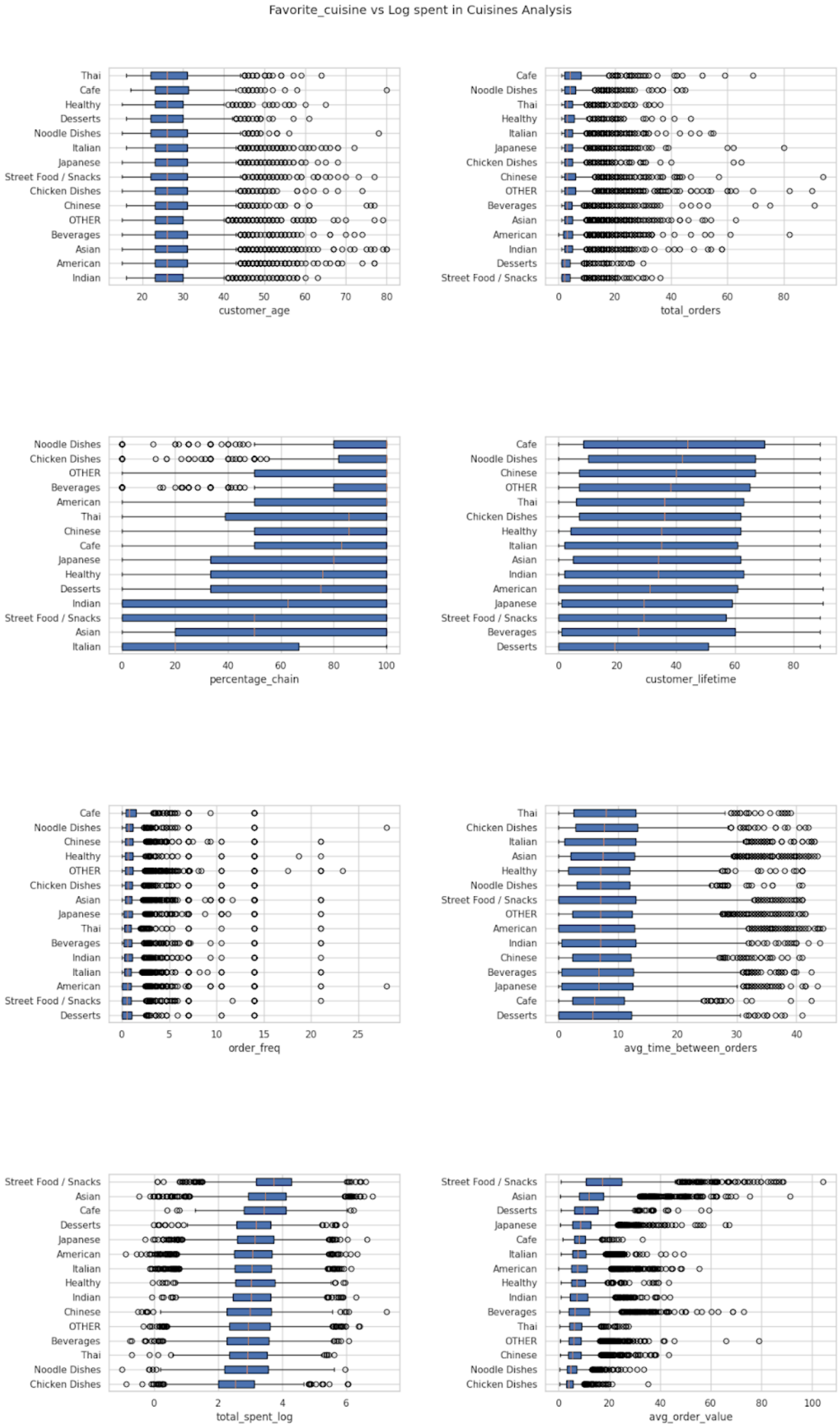

Figura 22 - Favorite cuisine vs. numerical variables.

Table 12 - Distribution of customer age and money spent on each cuisine by segmented region

1st numerical variable	Region	2nd numerical variable (Money spent on each cuisine)
Customer age	8	Centers 20-30: 2 centers (low/median and high spenders) Center 40-50: Chinese (high spenders)
	4	Center 30-40: Healthy (high spends ++) Center 50-60: Indian / Thai (median spenders)
	2	Center is usually lower than other regions Center 50-60: Beverages (median)

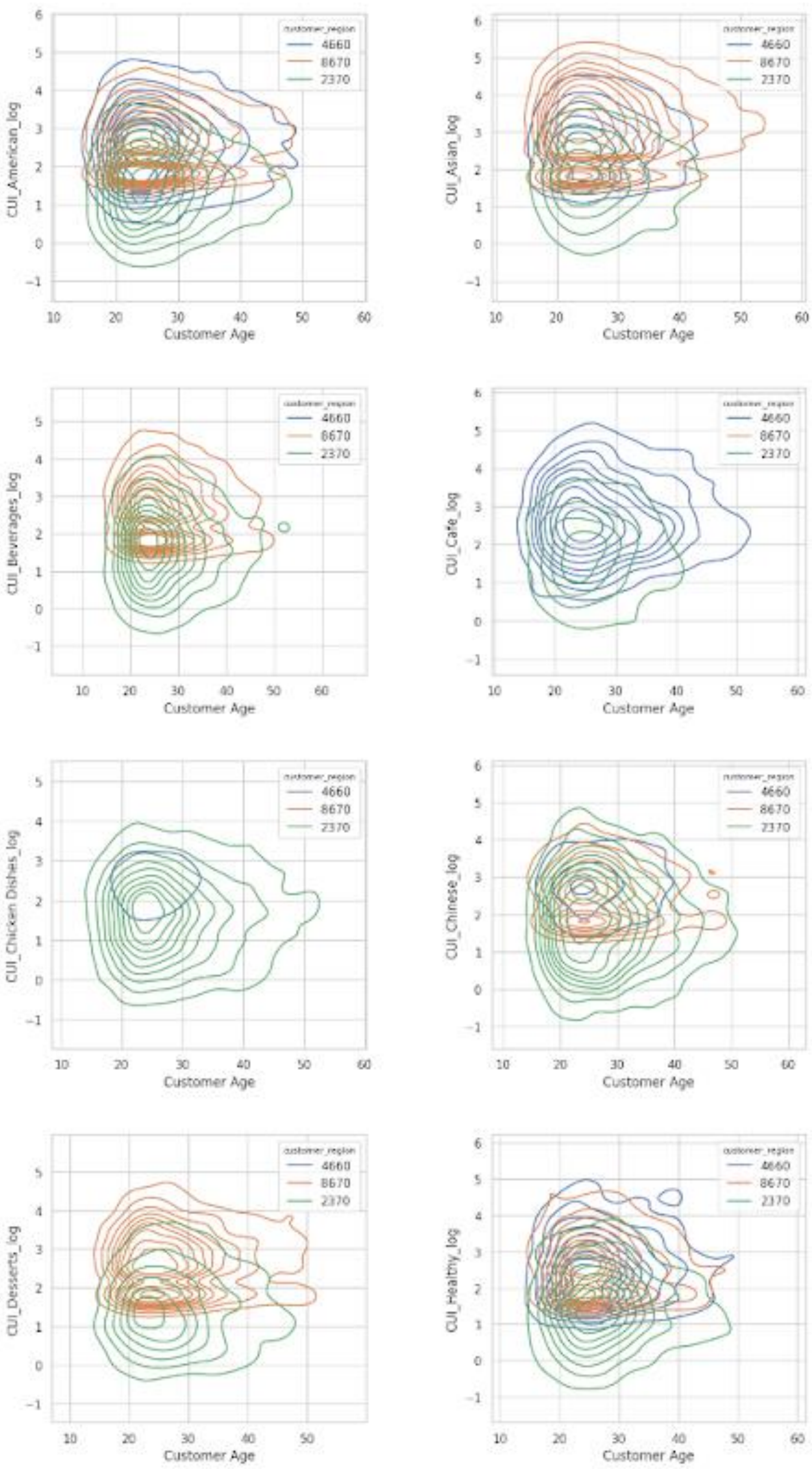

Figure 23.1 - Distribution of money spent per cuisine vs. customer age by region group.

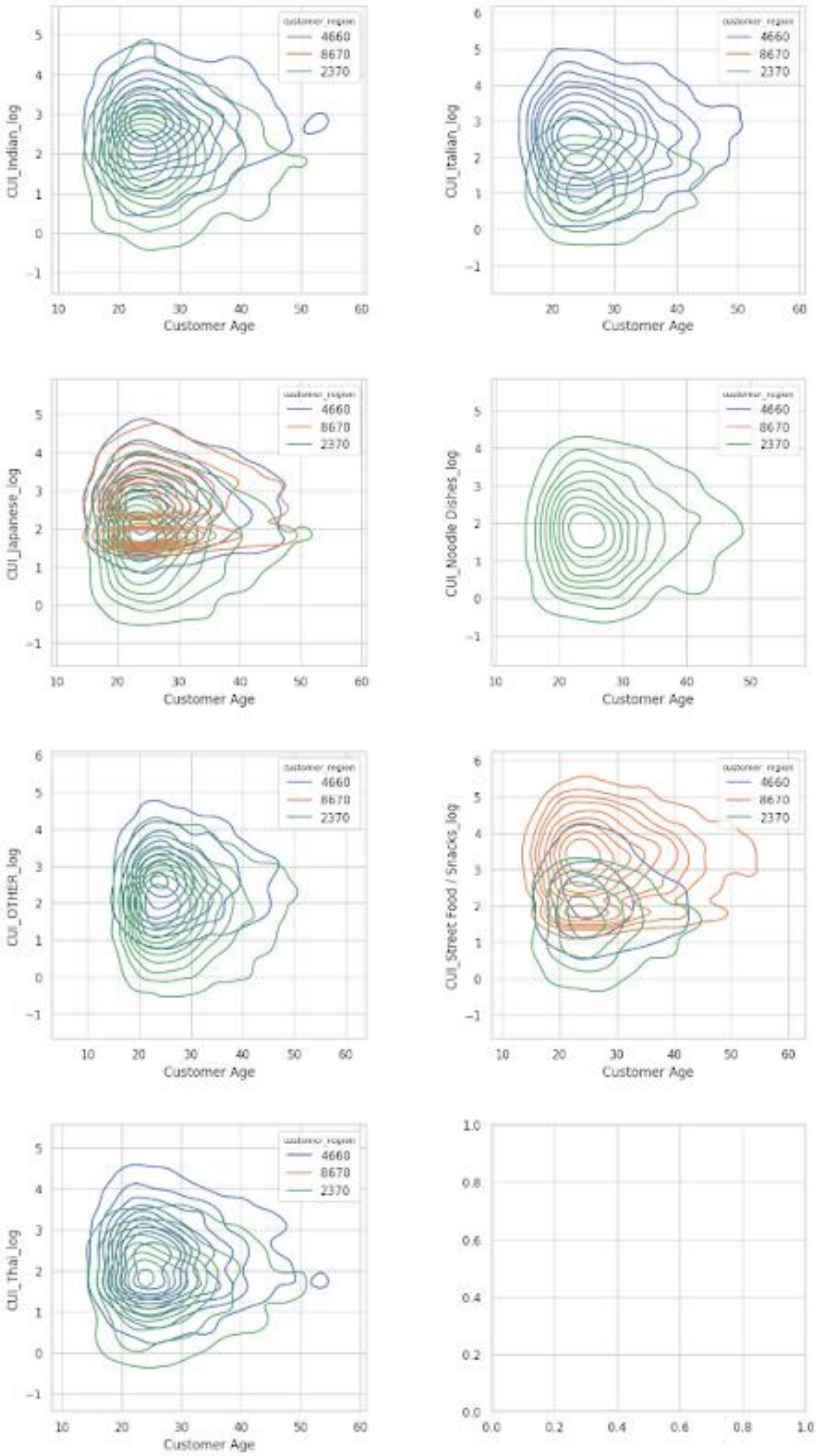

Figure 23.2 - Distribution of money spent per cuisine vs. customer age by region group.

Table 13 – Cuisines co-occurences (top 2)

Cuisine	Associated_Cuisine	Frequency
CUI_American	CUI_Asian	3789
	CUI_OTHER	2529
CUI_Asian	CUI_American	3789
	CUI_Japanese	2437
CUI_Beverages	CUI_Asian	2068
	CUI_American	1746
CUI_Cafe	CUI_American	607
	CUI_OTHER	571
CUI_Chicken Dishes	CUI_OTHER	1399
	CUI_American	1190
CUI_Chinese	CUI_OTHER	1249
	CUI_Asian	1225
CUI_Desserts	CUI_Asian	933
	CUI_American	680
CUI_Healthy	CUI_Asian	795
	CUI_American	703
CUI_Indian	CUI_American	1189
	CUI_OTHER	1150
CUI_Italian	CUI_American	2393
	CUI_OTHER	1799
CUI_Japanese	CUI_Asian	2437
	CUI_OTHER	2272
CUI_Noodle Dishes	CUI_OTHER	1012
	CUI_Chinese	746
CUI_OTHER	CUI_American	2529
	CUI_Asian	1859
CUI_Street Food/Snacks	CUI_Asian	2030
	CUI_American	1384
CUI_Thai	CUI_Asian	879
	CUI_American	872

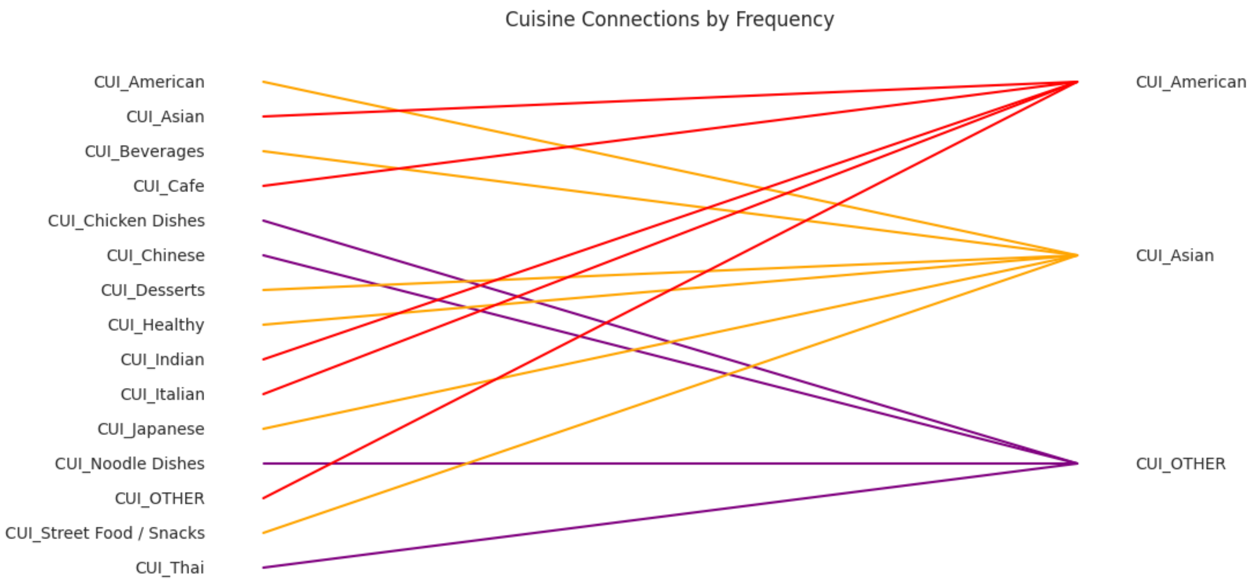

Figure 24 - Cuisine co-occurrences (top 1)

Table 14 - Weekday frequency co-occurrences (top 1)

dw1	dw2	Frequency
0	4	5345
1	4	5403
2	4	5664
3	4	5862
4	3	5862
5	6	6046
6	5	6046

Table 15 – Hour frequency co-occurrences (top 1)

hr1	h2	Frequency
0	1	258
1	3	297
2	3	451
3	4	704
4	3	704
5	4	612
6	10	485
7	10	562
8	9	1035
9	10	1913
10	11	2379
11	12	2509
12	11	2509
13	12	1925
14	15	1698
15	16	2459
16	17	2989
17	16	2989
18	17	2818
19	18	2146
20	19	1289
21	19	652
22	19	354
23	11	280