Consumer Perception Impacts on Olive Oil Consumption Choice: A Case Study Using Machine Learning Approach

Youmin Li

January 15, 2026

Abstract

Extra virgin olive oil (EVOO) and refined olive oil (ROO) are vertically differentiated healthy oil whose market outcomes depend heavily on consumer perceptions about sensory, price, and credibility of information cues. For agricultural food economic studies, rich survey instruments capture meaningful heterogeneity, but conventional empirical specifications can become sensitive to co-linearity and endogeneity. This study recognized how consumer perceptions shape olive oil consumption choices, combining unsupervised learning for data quality and structure detection with supervised models for prediction and inference. Using exploratory factor analysis to condense multi-item perception constructs, clustering to identify respondent archetypes and remove low-quality responses, and penalized models ensembles to predict consumption preference, we quantify (i) which perceptions matter most (EFA-PCA), (ii) how effects vary across consumer segments (Cluster and SEM), and (iii) out-of-sample performance under K-fold cross-validation. Results indicate that quality perceptions, taste imagery and price sensitivity are the dominant drivers, with heterogeneity across income and region.

Keywords: Olive Oil, Consumer Behavior, Agricultural Food, Machine Learning

1 Introduction

Olive oil markets are increasingly differentiated by perceived quality, health benefits, origin, certification, and sensory attributes [8, 2, 7]. Although traditional econometric studies often examine a small set of variables, modern surveys elicit dozens of Likert-scale items and product cues that are potentially collinear and noisy [4, 13]. This paper leverages machine learning (ML) to: (1) reduce dimensionality and denoise multi-item perception constructs; (2) identify latent consumer segments; and (3) predict consumption outcomes.

The purpose of the paper is to reveal the impacts of consumer perception on consumption choice using data from the choice experiment on olive oil consumption. So the study used factor and cluster analysis to find latent variables and incorporated consumer perception categorization into the Partial Least Squares Structural Equation Model (PLS-SEM) to explore moderation effects and mediation effects between consumer perception and consumption choice [14, 9]. To evaluate the prediction accuracy through cross validation. Then, the LASSO model was used to shrink corelation coefficients, and the prediction accuracy probability is greater than 90% [12]. This study contributes by offering an integrated pipeline that combines unsupervised learning for data curation and structure discovery with supervised learning for out-of-sample prediction and interpretable effect discovery.

2 Data

The cross-sectional data set in this paper were collected from consumers in Spain using a choice experiment and likert scale. The choice experimental questionnaire was developed through factor analysis of responses from a sample of 1031 adults. Perceived importance of olive oil attributes differs across individuals, and a five-class solution was estimated to describe each class in terms of knowledge and consumption of EVOO and the socio-demographic characteristics of the respondents. The data investigated the level of consumer awareness about the perceived health benefits, taste, price gap, and trust of EVOO and ROO by using likert scale. This information could be used to analyze consumer perceptions and preferences. Appendix table 8 briefly describes the core subjective factor variables in the data from that questionnaire. Consumers' trust perception and region attributes of the two olive oils are extrinsic attributes. Intrinsic attributes include perception of health benefits, taste preference, ingredient quality, and sensorial attributes. The data set source is from the Spanish Ministry of Agriculture and the University of Córdoba.

In addition, the data set includes the demographic information (age, gender, region, income, and Education level). The Table 1 shows a comparison between the sample and the population. The sample data is similar to the overall national socio-demographic background in terms of the age distribution, the gender ratio, and the distribution of education level. And the sample size of the used data set is 1031, so the study is somewhat representative of the national practical relevance of the results.

3 Methodology

3.1 Impacts pathway explanation

The effects of mediating and moderating variables are critical concerning explaining consumer behavior, and there is a gap in previous research [?]. To investigate the pathways linking perceptions to olive oil consumption, we estimated a partial least squares structural equation model (PLS-SEM) [6, 5], as shown in figure 1. The model included seven reflective latent constructs: taste perception (TASTE), price perception (PRICE), health perceptions (HEALTH_NEG), trust in extra virgin olive oil (TRUST_EVOO), trust in refined olive oil (TRUST_ROO), EVOO consumption (EVOO_CONS), and ROO consumption (ROO_CONS). Each construct was measured by multiple Likert-type items. During measurement refinement we dropped two indicators (PRICE_2 and NEG_CON_2) whose standardized loadings were below the conventional 0.70 threshold (see appendix), retaining only items with strong convergent validity. We analyze individual-level survey data on olive-oil consumption in Spain with information on EVOO and ROO consumption, perceptual items for Taste, Price, Trust (split by EVOO/ROO), and a retained single Health item; demographics include gender, income group, age group, and region (e.g., Andalucía, Comunidad de Madrid). Item wording and coding follow the thesis instrument. Likert items were numeric (higher = more favorable toward EVOO); negatively phrased taste items were reverse-coded.

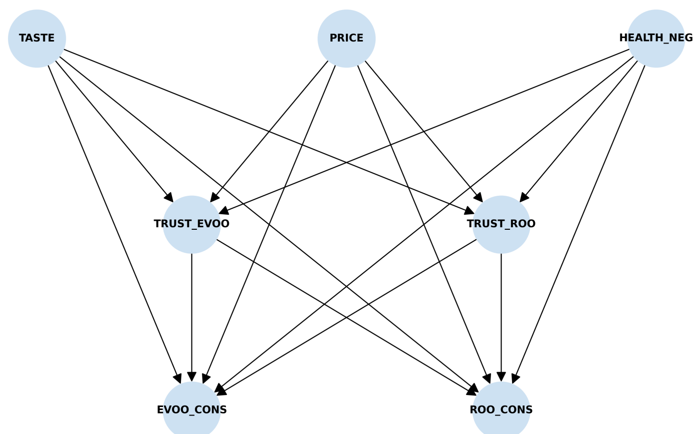


Figure 1: SEM Path Diagram showing perception - consumption pathways.

The final measurement model shows good reliability and validity. All retained indicators have loadings between about 0.78 and 0.95, and composite reliability (rhoC) exceeds 0.88 for all constructs (e.g., TASTE: rhoC = 0.86; TRUST_EVOO: rhoC = 0.89; TRUST_ROO: rhoC = 0.93; EVOO_CONS: rhoC = 0.92; ROO_CONS: rhoC = 0.95). Average variance extracted (AVE) is above 0.65 for every construct, indicating satisfactory convergent validity. HTMT ratios are comfortably below 0.90 for most construct pairs, supporting discriminant validity; the very high HTMT between EVOO_CONS and ROO_CONS reflects that many respondents consume both oils and that these two behavioral constructs are closely related rather than conceptually distinct attitudes.

After known about the impacts pathway, we adopt a two-stage approach. Stage I applies unsupervised learning to (a) clean the dataset, (b) summarize perception constructs, and (c) reveal latent consumer segments. Stage II fits supervised models, tuned via cross-validation, and evaluates predictive and structural importance.

3.2 Unsupervised machine learning for Structure & Data Quality

Factor Analysis (EFA) - PCA

To reduce dimensionality and uncover latent constructs among the 1031 survey respondents,We apply exploratory factor analysis (EFA) using the principal component analysis method (PCA) on standardized perception items to extract low-dimensional latent constructs. We select 5 by eigenvalue > 1 criteria, and oblimin rotation was performed on all perception items, including trust, taste, price, and health indicators for both extra virgin (EVOO) and refined olive oil (ROO). Factor scores $\hat{\mathbf{f}}_i$ are used as inputs to supervised models. Sampling adequacy was high (KMO = 0.823) and Bartlett's test of sphericity was highly significant ($\chi^2 = 7136.94$, $p < 0.001$) (appendix table 4), confirming that the inter-item correlations were sufficient for factor analysis. This hybrid approach allows for a better understanding of the underlying structure of the data while reducing the dimensionality for further analysis of structural visualization.

Consumer segment cluster

Following the extraction of five latent perception factors (F1–F5), we conducted unsupervised segmentation using k-means clustering on the standardized factor scores [12]. The optimal number of clusters was selected using the average silhouette coefficient. As shown in appendix figure 6, silhouette widths peaked at $k = 5$, indicating that a five cluster solution provides the best-defined separation between consumer segments. Tthe factor-score centroids reveal five distinct perception-based consumer segments, presented in table 1. These segments differ systematically by their levels of trust (EVOO and ROO), taste valuation, price and health sensitivity, and consumption orientation. The clusters can be interpreted as in appendix 6. The appendix figure 5 plot visualized the clear directional differences among segments, especially along the consumption polarity (PC1) and taste/trust dimensions (PC2), further confirming that these five clusters capture meaningful heterogeneity in consumer perceptions.

Table 1: Mean standardized factor scores by perception cluster

Cluster	F1	F2	F3	F4	F5
1 (EVOO favor)	−1.35	0.42	−0.74	0.40	−1.05
2 (Mild flavor)	0.00	0.09	−0.30	−0.83	−0.45
3 (Price sensitive)	0.85	0.17	0.95	0.03	0.75
4 (Taste driven)	0.30	0.51	−0.51	0.75	0.28
5 (ROO favor)	0.10	−1.81	0.83	−0.37	0.60

3.3 Supervised machine learning

3.3.1 Penalized Generalized Linear Models (LASSO)

For binary outcomes (consumption choice on *ROO* or *EVOO*), we fit LASSO-penalized logistic regression:

$$\Pr(Y_i = 1 \mid \mathbf{z}_i) = \sigma(\beta_0 + \mathbf{z}_i^\top \boldsymbol{\beta}), \quad \sigma(t) = \frac{1}{1 + e^{-t}} \tag{1}$$

$$\hat{\boldsymbol{\beta}}(\lambda) = \arg\min_{\boldsymbol{\beta}} \left\{ -\ell(\boldsymbol{\beta}) + \lambda \|\boldsymbol{\beta}\|_1 \right\}, \tag{2}$$

where $\mathbf{z}_i$ includes factor scores, price cues, and demographics. The L1 penalty induces sparsity, improving interpretability when many correlated inputs exist.

We fit a logistic regression with a penalty (LASSO) to predict the binary outcome of EVOO choice. The LASSO penalty performs feature selection by shrinking less-informative coefficients to zero [11]. Importantly, even highly flexible models (RF, XGBoost, GAM) do not outperform logistic regression, indicating that the structure of the predictive relationship is largely linear and additive, with little evidence of deeper nonlinear interactions [12]. This provides strong evidence that the extracted perceptual factors are stable and interpretable predictors of olive-oil choice, supporting their use for segmentation and subsequent policy or marketing analysis.

3.3.2 K-fold Cross-Validation (CV).

The data are partitioned into 10 folds; each fold serves once as a validation set while the model is trained on the remaining $K-1$ folds. We select $\lambda_{\min}$ with the lowest mean validation loss (e.g., log-loss or Brier score) and optionally λ_{1SE} for a more parsimonious solution. The results show that consumer choice between EVOO and ROO is well-explained by a low-dimensional set of perception constructs (trust, taste, price sensitivity, health beliefs). We fixed λ at that value and re-trained the penalized logit model on the entire training set (80% of data stratified by outcome). The final model was then evaluated on the 20% hold-out test set. We assessed discrimination using the ROC AUC and accuracy at a 0.5 probability threshold, and calibration via log-loss and bias. The bias was measured as the mean predicted probability minus the actual fraction of EVOO users in the test set [11]. A confusion matrix was also compiled to inspect classification errors.

4 Results

4.1 Cross-Validated Performance

We compare models using repeated CV metrics (AUC, accuracy, log-loss for classification), calibration (reliability curves figure 2), and fairness checks across observed subgroups. Table 2 reports the predictive performance of five classification models—standard logistic regression, LASSO-penalized logistic regression, Random Forest, XGBoost boosting, and a GAM-logit—using 10-fold cross-validation to evaluate out-of-sample performance. The model showed high discrimination (AUC = 0.91) and good accuracy (87%). Calibration was good as indicated by a low log-loss of 0.33. The mean predicted probability of choosing EVOO was within 2% points of the observed rate (indicating only slight bias).

Table 2: Out-of-sample performance (10-fold CV). Best in bold.

Model	**ROC_AUC**	**Accuracy**	**Log-loss**	**Notes**
Logistic (baseline)	0.90	0.86	0.33	–
LASSO-logit	**0.91**	**0.87**	**0.33**	$\lambda_{\min}$/1SE
Random Forest	0.88	0.85	0.35	500 trees
Boosting	0.88	0.84	0.35	tuned
GAM-logit	0.89	0.87	0.32	EDF summary

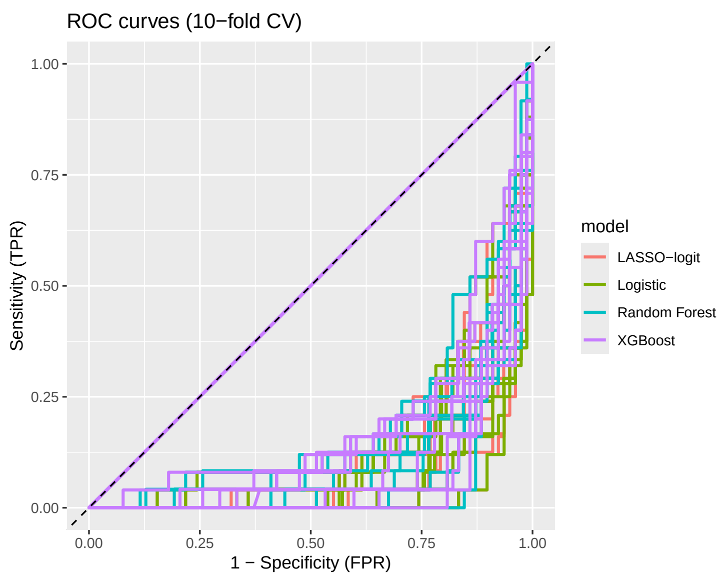


Figure 2: ROC curves (10-fold CV) for the binary task: prefers EVOO (=1).
Note: Baseline: OLS (continuous) and logistic regression (binary). Shrinkage: LASSO via cross-validated penalty. Flexible Semi-parametric: GAM with thin-plate splines/smooths for key factors (e.g., price gap, taste, trust). Tree Ensembles: Random forest and gradient boosting. Validation: Stratified K-fold CV ($K = 10$). For classification, report AUC, accuracy, log-loss, and calibration.

4.2 LASSO-Logit regression model for prediction

Finally, table 3 displays the LASSO-logit selected coefficient estimates for each predictor (all features were retained at the chosen λ) along with their exponentiated odds ratios. Positive coefficients indicate higher likelihood of choosing EVOO. The LASSO-logit model selected both perceptual and demographic predictors of EVOO choice.

Among the perception factors, consumption polarity (F1), trust in ROO (F2), taste (F3), and trust in EVOO (F4) were retained, while the health/price factor (F5) was shrunk to zero. F1 and F3 show strong negative effects, indicating that ROO-oriented usage patterns and taste-driven preferences substantially reduce the likelihood of choosing EVOO. In contrast, trust in ROO (F1) and trust in EVOO (F2) post opposite effects on the odds of choosing EVOO, with odds ratios of 0.29 and 2.22, respectively. Demographic variables were mostly shrunk toward zero, suggesting limited predictive value once perceptions are controlled for. Income, discrete age grouping, and residence in producing areas remained in the model but with small effect sizes (OR between 1.09 and 1.18). We see that living in an olive-oil producing area is associated with increased odds of EVOO use (OR=1.18), whereas a one unit higher "Price-sensitive" factor score (Factor 3) lowers the odds (OR=0.4). Overall, the selected predictors and their odds ratios align with domain expectations.

Table 3: LASSO-logit coefficients and odds ratios for predicting EVOO choice

(Positive β increases odds of EVOO choice.)

Factor	**Coefficient (β)**	**Odds Ratio**
F1: Trust in ROO	-1.13	0.29
F2: Trust in EVOO	0.92	2.22
F3: Price perception	-0.86	0.40
F4: Taste perception	0.60	1.73
F5: Health perceptiony	0.00	1.00
gender	0	1
income	0.09	1.1
age_discret	0.09	1.09
education	0.04	1.04
region	0.17	1.18

Appendix table 7 provides the confusion matrix at the 0.5 threshold, showing that most

EVOO choosers were correctly identified (only 50 of 800 were missed) while about half of non-choosers were misclassified as EVOO users. This reflects EVOO's high prevalence and the model's tendency to predict "1" for most cases. The confusion matrix shows that the model correctly classified 94.2% of EVOO users (TP = 147, FN = 9) and 65.3% of ROO users (TN = 32, FP = 17). Figure 3 visualizes predicted probabilities for both the training and test sets. EVOO choice (1) cluster around high predicted probabilities, whereas ROO choice (0) cluster around lower predicted probabilities, with similar separation patterns in both datasets. The close correspondence between training and test distributions indicates that the LASSO-logit model generalizes well and exhibits limited overfitting. The LASSO-logit model achieved strong predictive performance on the test set. The mean prediction bias was extremely small (Bias = –0.0024), indicating that the model's predicted probabilities are well-calibrated, consistent with the ROC–AUC of 0.91. Most mis-classification occurred among ROO users, who are underrepresented in the sample, but the model performed exceptionally well for the dominant EVOO user class.

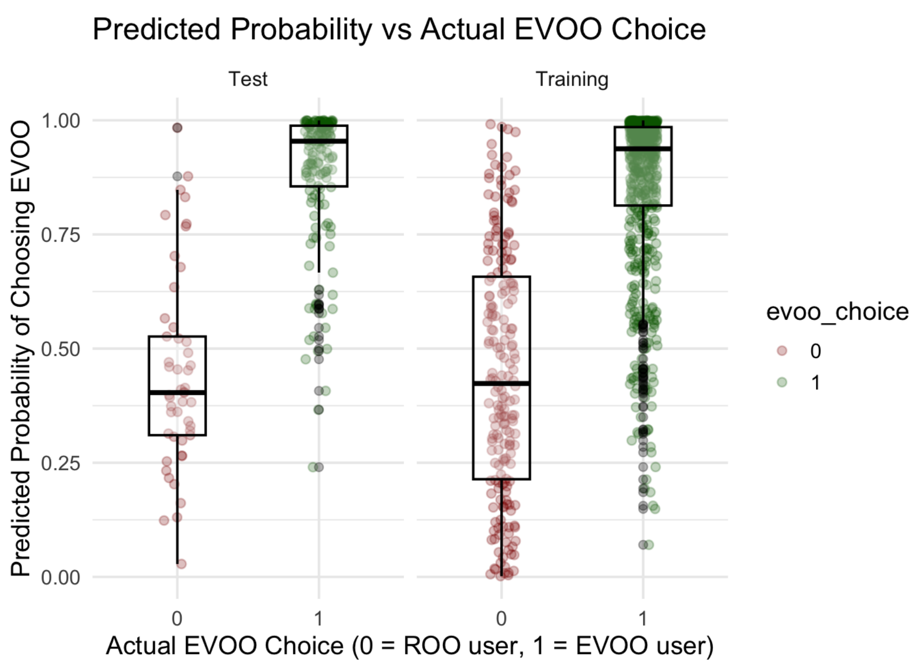


Figure 3: LASOO-logit model predict probability on EVOO-choice.

5 Discussion and Conclusions

This study has considered both olive oil attributes and consumer perception of trust, taste, price, and socio-demographic characteristics to use an ML framework linking rich perception data to olive oil consumption choices. LASSO consistently highlighted trust perceptions, taste imagery, and region identification as primary drivers, with notable heterogeneity across consumer segments. People in olive oil producing regions pay more attention to the perception of taste. Perceived trust is the primary influence factor on most consumer groups. And eople who prioritize trust are more likely to be aware of the health benefits associated with olive oil. Many people believe that any kind of olive oil is a healthy food in itself compared to other foods, so the effect of health perception is not a significant variable compared to other influencing factors [10].

By employing novel machine learning approaches and results illustration, this paper offers the following implications for policy, market development and academic. Targeted messaging tailored to segment-specific salience (health vs. price vs. sensory); and product positioning that aligns sensory expectations with verified quality claims are necessary. Future work could integrate consumer segments and demographic characters, conduct multiple group analysis in casual empirical studies to reveal the differences among different groups, and examine substitution patterns among edible oils. Accurate and effective scientific knowledge of olive oil helps raise consumers' health awareness and WTP of quality olive oil. More importantly, eco-friendly food labeling should be promoted to raise consumers' awareness of ecological services when making consumption choices [1]. To improve the sales of EVOO and maintain the competitiveness of Spanish olive oil in the international market, it is important to focus on consumer trust perception and healthy knowledge on EVOO. Different types of olive oil should be marketed to different consumer groups with adapted marketing strategies [3].

References

[1] Georgios Banias, Charisios Achillas, Christos Vlachokostas, Nicolas Moussiopoulos, and Maria Stefanou. Environmental impacts in the life cycle of olive oil: a literature review. *Journal of the Science of Food and Agriculture*, 97(6):1686–1697, 2017.

[2] Sofía Boza, Aracely Núñez-Mejía, Marcos Mora, and Dorotea López. Determining factors of the international competitiveness of Extra- Virgin Olive Oil (EVOO) from Spain and Chile. *International Association of Agricultural Economists*, 2023.

[3] Karen Brunsø, Thomas Ahle Fjord, and Klaus G Grunert. CONSUMERS' FOOD CHOICE AND QUALITY PERCEPTION.

[4] Vincenzina Caputo and Jayson L. Lusk. The Basket-Based Choice Experiment: A Method for Food Demand Policy Analysis. *Food Policy*, 109:102252, May 2022.

[5] Wynne W. Chin. Bootstrap Cross-Validation Indices for PLS Path Model Assessment. In Vincenzo Esposito Vinzi, Wynne W. Chin, Jörg Henseler, and Huiwen Wang, editors, *Handbook of Partial Least Squares: Concepts, Methods and Applications*, pages 83–97. Springer, Berlin, Heidelberg, 2010.

[6] Klaus G. Grunert. Food quality and safety: consumer perception and demand. *European Review of Agricultural Economics*, 32(3):369–391, September 2005.

[7] Youmin Li. Graduate School of Agricultural and Rural Development Seoul National University Agricultural and Resource Economics Major. Master's thesis, Seoul National University, South Korea, 2024.

[8] Nieves Martínez, Isabel Prieto, Marina Hidalgo, Ana Belén Segarra, Ana M. Martínez-Rodríguez, Antonio Cobo, Manuel Ramírez, Antonio Gálvez, and Magdalena Martínez-Cañamero. Refined versus Extra Virgin Olive Oil High-Fat Diet Impact on Intestinal

Microbiota of Mice and Its Relation to Different Physiological Variables. *Microorganisms*, 7(2):61, February 2019. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

[9] Mehmet Mehmetoglu and Sergio Venturini. *Structural Equation Modelling with Partial Least Squares Using Stata and R.* February 2021.

[10] Cristina Nocella, Vittoria Cammisotto, Luca Fianchini, Alessandra D'Amico, Marta Novo, Valentina Castellani, Lucia Stefanini, Francesco Violi, and Roberto Carnevale. Extra Virgin Olive Oil and Cardiovascular Diseases: Benefits for Human Health. *Endocrine, Metabolic & Immune Disorders - Drug Targets(Formerly Current Drug Targets - Immune, Endocrine & Metabolic Disorders)*, 18(1):4–13, January 2018.

[11] Tomoyuki Obuchi and Yoshiyuki Kabashima. Cross validation in LASSO and its acceleration. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(5):053304, May 2016. Publisher: IOP Publishing and SISSA.

[12] Fariha Sohil, Muhammad Umair Sohali, and Javid Shabbir. An introduction to statistical learning with applications in R: by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, New York, Springer Science and Business Media, 2013, $41.98, eISBN: 978-1-4614-7137-7. *Statistical Theory and Related Fields*, 6(1):87–87, January 2022.

[13] Andrew Steptoe, Tessa M. Pollard, and Jane Wardle. Development of a Measure of the Motives Underlying the Selection of Food: the Food Choice Questionnaire. *Appetite*, 25(3):267–284, December 1995.

[14] Sergio Venturini and Mehmet Mehmetoglu. plssem: A Stata Package for Structural Equation Modeling with Partial Least Squares. *Journal of Statistical Software*, 88, January 2019.

A Appendix - Additional Tables and Figures

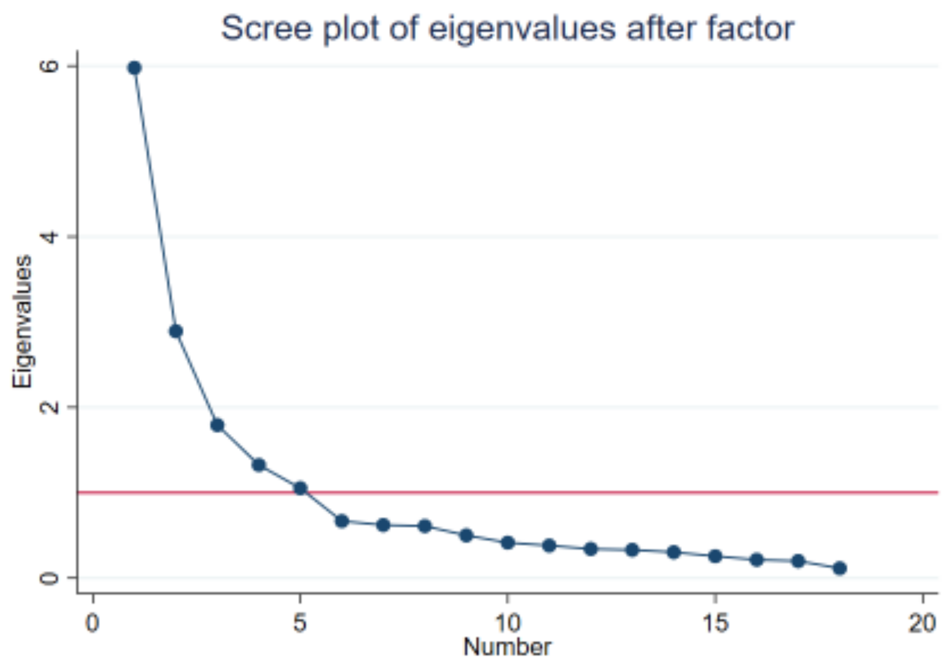


Figure 4: Screen Plot

Table 4: Bartlett's test of sphericity and Kaiser–Meyer–Olkin (KMO).

Statistic	**Value**	**Notes**
Bartlett's test of sphericity:		
Chi-square (χ^2)	10,419.814	
Degrees of freedom	153	
p-value	0.000	variables are not intercorrelated
Kaiser–Meyer–Olkin (KMO)	0.8799	Sampling adequacy (meritorious)

Table 5: Factor rotation matrix from principal component extraction.

	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**
Factor 1	0.8839	0.6300	0.6981	-0.5602	-0.2667
Factor 2	-0.0694	0.6872	0.0250	0.6798	0.2853
Factor 3	0.3315	-0.3551	0.2712	0.4350	-0.0179
Factor 4	-0.0495	-0.0689	0.3182	-0.1862	0.8918
Factor 5	-0.3187	0.0049	0.5807	0.0156	-0.2277

Table 6: Summary of factor interpretation based on rotated loadings

Factor	**Interpretation / Dominant items**
F1	Trust in ROO
F2	Trust in EVOO
F3	Price perception
F4	Taste perception
F5	Health perception

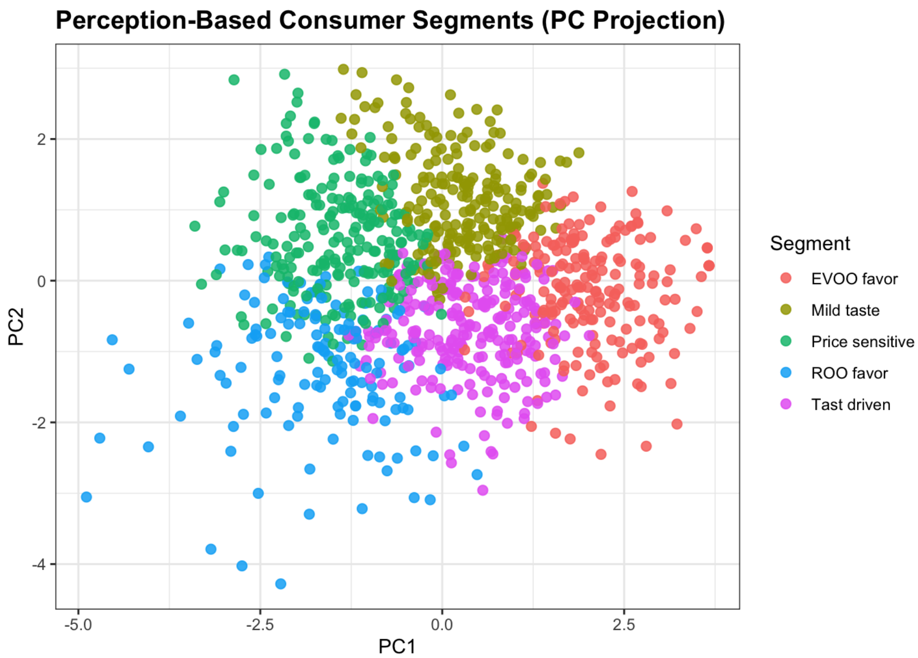


Figure 5: Perception-based consumer segments (PC projection)

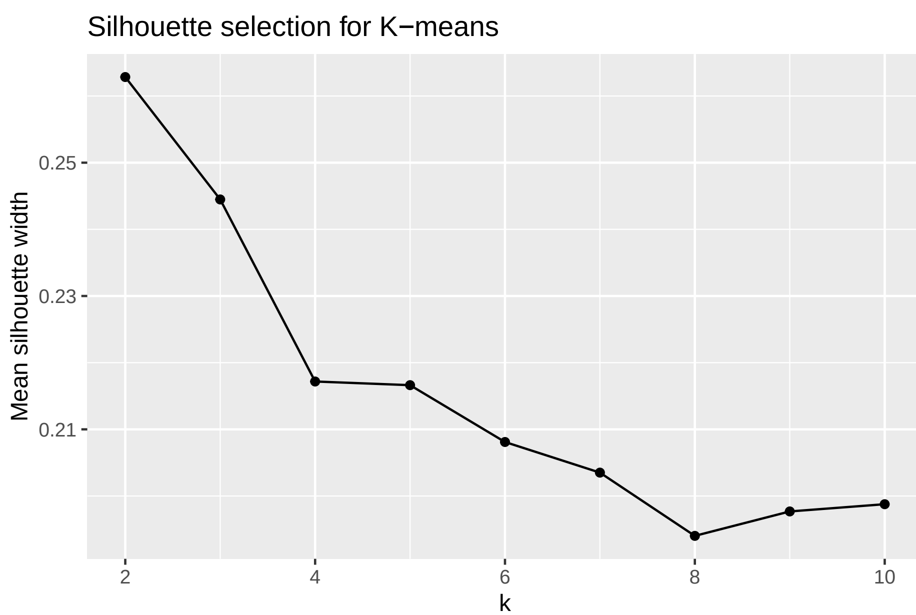
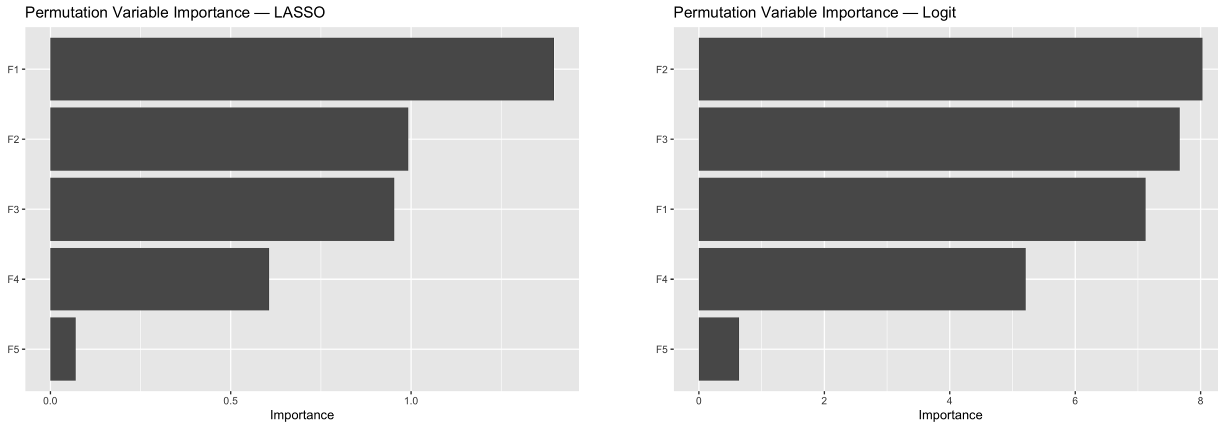


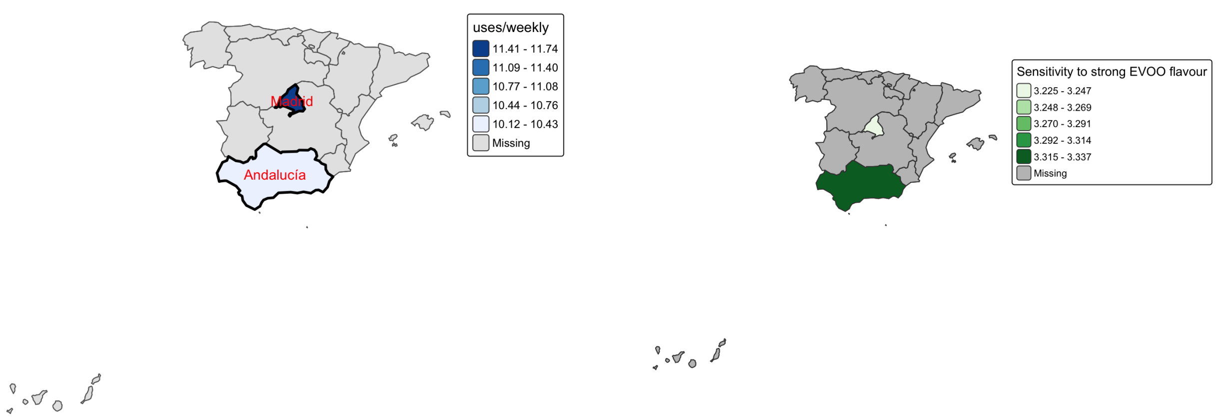
Figure 6: Cluster segments silhouette



(a) LASSO (b) Logit

Figure 7: Permutation variable importance (LASSO / Logit)

Note: Variable importance (permutation), partial dependence/ALE.



(a) Uses per week per household (b) Prefer to strong flavor

Figure 8: Map of difference of Madrid and Andalucía

Table 7: Confusion matrix for LASSO-logit predictions (test set)

	Actual ROO (0)	**Actual EVOO (1)**
Predicted ROO (0)	32	9
Predicted EVOO (1)	17	147

B Data Description - metadata

Table 8: Description of variables used in the olive oil consumption study.

Variable	Description
EVOO Use	Weekly uses of extra virgin olive oil (EVOO) in the household, measured as number of meal uses per week.
ROO Use	Weekly uses of refined olive oil (ROO) in the household, measured as number of meal uses per week.
EVOO Con	Monthly consumption of EVOO in the household, in liters per month per person.
ROO Con	Monthly consumption of ROO in the household, in liters per month per person.
Trust ROO1	Your degree of trust in ROO, 1–7 Likert scale.
Trust ROO2	The degree to which you need ROO, 1–7 Likert scale.
Trust ROO3	The degree to which you recommend ROO, 1–7 Likert scale.
Trust ROO4	The perceived quality of ROO, 1–7 Likert scale.
Trust EVOO1	Your degree of trust in EVOO, 1–7 Likert scale.
Trust EVOO2	The degree to which you need EVOO, 1–7 Likert scale.
Trust EVOO3	The degree to which you recommend EVOO, 1–7 Likert scale.
Trust EVOO4	The perceived quality of EVOO, 1–7 Likert scale.
Taste 1	"EVOO is less versatile in the kitchen due to its taste" (reverse-coded), 1–7 Likert scale.
Taste 2	"I prefer a mild and light-flavored olive oil" (reverse-coded), 1–7 Likert scale.
Taste 3	"EVOO adds too strong a flavor for most dishes" (reverse-coded), 1–7 Likert scale.
Price 1	"Extra virgin olive oil has a suitable price," 1–7 Likert scale.
Price 2	"The price gap between EVOO and ROO is small," 1–7 Likert scale.
Price 3	"Considering annual food outlay, EVOO is a cheap product," 1–7 Likert scale.
Price 4	"Considering its features, ROO has a high price," 1–7 Likert scale.
Health 1	"EVOO and ROO have the same health benefits," 1–7 Likert scale.
Health 2	"EVOO is not as good for frying as ROO," 1–7 Likert scale.
Health 3	"EVOO and ROO have the same features except for taste," 1–7 Likert scale.

C Additional information for PLS-SEM

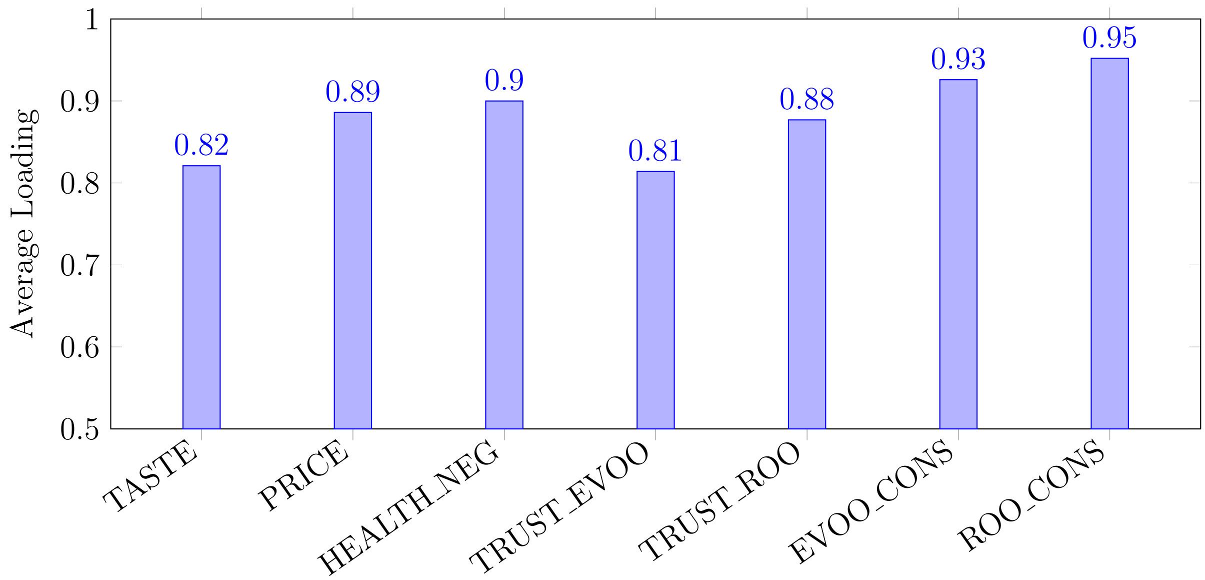


Figure 9: Average standardized indicator loadings for each reflective construct.

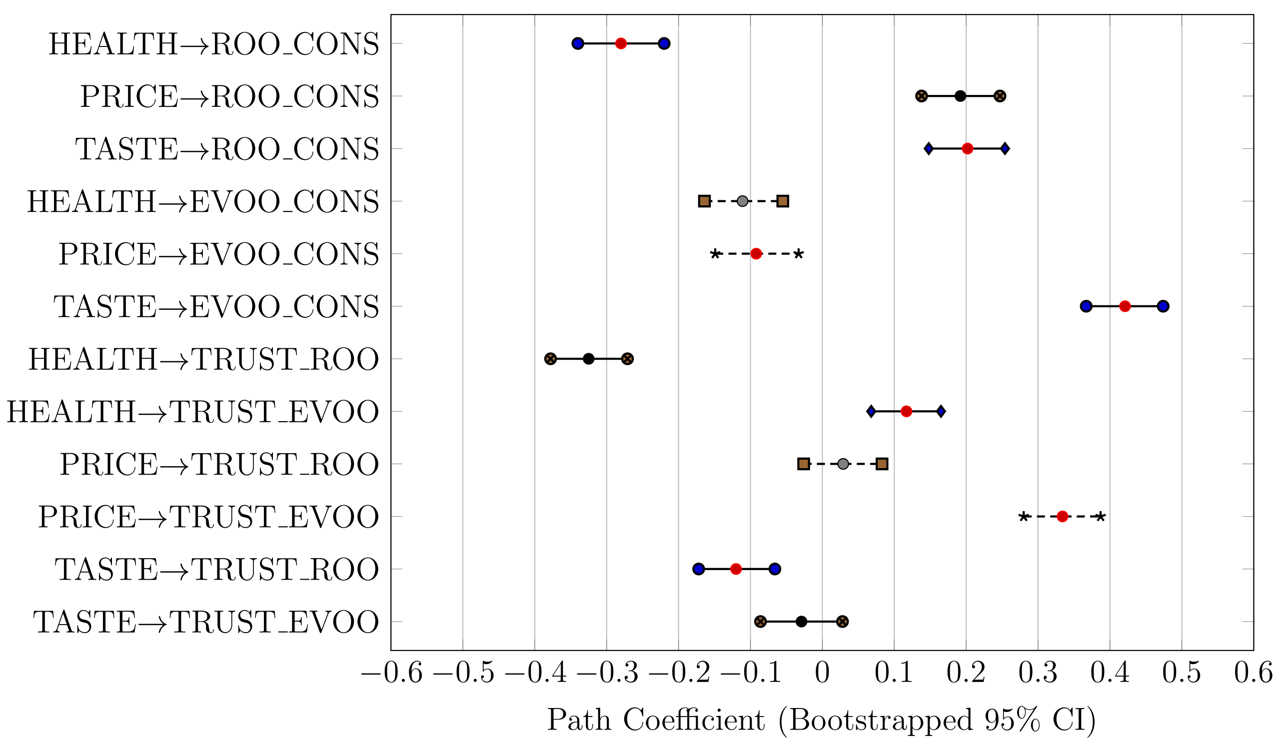


Figure 10: Forest plot of structural path estimates with 95% bootstrap CIs.

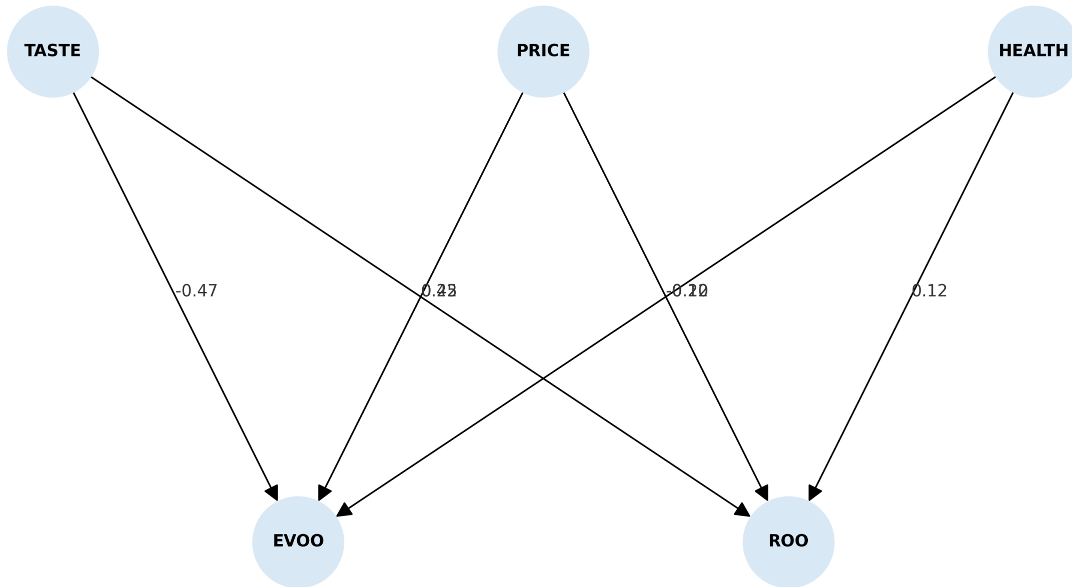


Figure 11: Total effects (direct + mediated) from key perceptions to EVOO and ROO consumption.

Table 9: Bootstrapped Total Effects (Direct + Indirect)

Total Effect	**Est.**	**Mean**	**SD**	**95% CI**
TASTE → EVOO_CONS	-0.471	-0.471	0.027	[-0.523, -0.416]
TASTE → ROO_CONS	0.454	0.453	0.026	[0.401, 0.504]
PRICE → EVOO_CONS	0.215	0.215	0.026	[0.165, 0.267]
PRICE → ROO_CONS	-0.198	-0.198	0.027	[-0.250, -0.145]
HEALTH_NEG → EVOO_CONS	-0.117	-0.117	0.029	[-0.175, -0.059]
HEALTH_NEG → ROO_CONS	0.119	0.119	0.028	[0.064, 0.173]
TRUST_EVOO → ROO_CONS	-0.221	-0.222	0.030	[-0.281, -0.161]
TRUST_ROO → EVOO_CONS	-0.273	-0.273	0.024	[-0.320, -0.226]

Table 10: Measurement Model: Loadings, Reliability, and AVE

Construct	**Indicator**	**Loading**	**CR (rhoC)**	**AVE**
TASTE	taste_1	0.804		
	taste_2	0.788	0.862	0.675
	taste_3	0.871		
PRICE	price_1	0.933	0.881	0.787
	price_3	0.839		
HEALTH_NEG	neg_con_1	0.909	0.895	0.811
	neg_con_3	0.891		
TRUST_EVOO	act_evoo_1	0.746		
	act_evoo_2	0.838	0.888	0.664
	act_evoo_3	0.859		
	act_evoo_4	0.812		
TRUST_ROO	act_roo_1	0.872		
	act_roo_2	0.869	0.930	0.768
	act_roo_3	0.886		
	act_roo_4	0.881		
EVOO_CONS	evoo_con	0.928	0.923	0.856
	evoo_uses	0.923		
ROO_CONS	roo_con	0.954	0.951	0.906
	roo_uses	0.950		

Table 11: HTMT Discriminant Validity

Construct Pair	**HTMT**	**Mean**	**SD**	**95% CI**
TASTE – PRICE	0.093	0.100	0.037	[0.039, 0.181]
TASTE – HEALTH_NEG	0.515	0.515	0.037	[0.439, 0.588]
TASTE – TRUST_EVOO	0.411	0.410	0.035	[0.340, 0.479]
TASTE – TRUST_ROO	0.439	0.439	0.031	[0.378, 0.500]
TASTE – EVOO_CONS	0.669	0.669	0.029	[0.612, 0.724]
TASTE – ROO_CONS	0.624	0.624	0.028	[0.567, 0.677]
PRICE – HEALTH_NEG	0.048	0.070	0.027	[0.025, 0.134]
PRICE – TRUST_EVOO	0.271	0.271	0.034	[0.203, 0.336]
PRICE – TRUST_ROO	0.136	0.136	0.039	[0.061, 0.213]
PRICE – EVOO_CONS	0.301	0.301	0.038	[0.226, 0.374]
PRICE – ROO_CONS	0.272	0.272	0.037	[0.197, 0.344]
HEALTH_NEG – TRUST_EVOO	0.243	0.242	0.037	[0.169, 0.314]
HEALTH_NEG – TRUST_ROO	0.593	0.593	0.027	[0.539, 0.646]
HEALTH_NEG – EVOO_CONS	0.369	0.370	0.037	[0.296, 0.442]
HEALTH_NEG – ROO_CONS	0.350	0.350	0.035	[0.283, 0.418]
TRUST_EVOO – TRUST_ROO	0.093	0.100	0.012	[0.077, 0.127]
TRUST_EVOO – EVOO_CONS	0.557	0.557	0.031	[0.498, 0.615]
TRUST_EVOO – ROO_CONS	0.412	0.412	0.035	[0.342, 0.480]
TRUST_ROO – EVOO_CONS	0.460	0.460	0.026	[0.411, 0.510]
TRUST_ROO – ROO_CONS	0.476	0.476	0.023	[0.432, 0.520]
EVOO_CONS – ROO_CONS	0.985	0.986	0.010	[0.966, 1.005]

Table 12: Bootstrapped Structural Paths (5,000 resamples)

Path	Est.	Mean	SD	t-value	95% CI
TASTE → TRUST_EVOO	-0.280	-0.281	0.031	-9.18	[-0.340, -0.220]
TASTE → TRUST_ROO	0.192	0.192	0.028	6.86	[0.138, 0.247]
TASTE → EVOO_CONS	-0.325	-0.325	0.027	-12.05	[-0.378, -0.271]
TASTE → ROO_CONS	0.334	0.333	0.027	12.31	[0.280, 0.387]
PRICE → TRUST_EVOO	0.202	0.202	0.028	7.30	[0.148, 0.254]
PRICE → TRUST_ROO	-0.111	-0.111	0.028	-3.96	[-0.164, -0.055]
PRICE → EVOO_CONS	0.117	0.117	0.025	4.69	[0.068, 0.165]
PRICE → ROO_CONS	-0.120	-0.120	0.026	-4.54	[-0.172, -0.066]
HEALTH_NEG → TRUST_EVOO	-0.092	-0.092	0.030	-3.10	[-0.149, -0.033]
HEALTH_NEG → TRUST_ROO	0.421	0.421	0.028	15.27	[0.367, 0.474]
HEALTH_NEG → EVOO_CONS	0.029	0.029	0.028	1.05	[-0.026, 0.083]
HEALTH_NEG → ROO_CONS	-0.029	-0.028	0.029	-1.00	[-0.086, 0.028]
TRUST_EVOO → EVOO_CONS	0.334	0.334	0.027	12.45	[0.281, 0.386]
TRUST_EVOO → ROO_CONS	-0.221	-0.222	0.030	-7.28	[-0.281, -0.161]
TRUST_ROO → EVOO_CONS	-0.273	-0.273	0.024	-11.22	[-0.320, -0.226]
TRUST_ROO → ROO_CONS	0.302	0.302	0.025	12.31	[0.254, 0.350]

D Multi-Group Analysis (MGA)

To assess heterogeneity, we conduct MGA by estimating models within predefined groups (e.g., income tertiles, frequent vs. infrequent users, age groups) and testing differences in key effects or importances. For GLMs, we compare coefficients across groups using bootstrap standard errors and permutation tests on cross-validated losses. For RF, we compare group-specific permutation importances and PD/ALE shapes. Formally, for a parameter/importance $\theta^{(g)}$ in group g, we test

$$H_0 : \theta^{(g_1)} = \theta^{(g_2)} \quad \text{vs.} \quad H_A : \theta^{(g_1)} \neq \theta^{(g_2)}, \tag{3}$$

using a null distribution from label permutations or pooled bootstrap resampling.

Compare key effects/importances across subgroups. Example: Quality perception is more predictive among high-income households (importance difference = 0.07, perm. $p = 0.02$), whereas price sensitivity dominates for low-income groups. Provide a table of group-specific metrics and a plot of PD curves by group.

E Data and Code Share

- **Github**: `https://github.com/liyoumin/Machine-learning-project`
- **Presentation record**: `https://liyoumin.github.io/personalweb/teaching/js/`
- **Cross-validation**: stratified K-fold for classification; repeated CV for stability.