Phoenix: International Multidisciplinary Research Journal
Vol 4, No. 3, July-Sep, 2026
ISSN: 2583-6897

CONSUMER DECISION-MAKING IN THE ORGANIC FOOD MARKET: A DEMOGRAPHIC PERSPECTIVE ON GENDER, MARITAL STATUS, AND LOCATION

Dr. Rakesh Kumar Upadhyay
Assistant professor, Department of B.Com (Hons.)
FMS, VBSPU Jaunpur

Anupam Kumar
Research Scholar, MBA, VBSPU Jaunpur

Abstract— Organic food consumption continues to expand across emerging economies, yet the demographic structure of this demand remains insufficiently quantified at the intersection of gender, marital status, and geographic location. This study surveys 850 consumers across urban, semi-urban, and rural strata and applies a mixed quantitative pipeline combining chi-square association tests, one-way and two-way analysis of variance (ANOVA), binary logistic regression, and a Random Forest ensemble classifier to model regular organic purchase behavior and willingness-to-pay (WTP) premiums. Results show a statistically significant association between location and purchase regularity ($\chi^2=111.52$, $p<0.001$, Cramer's $V=0.362$), a significant gender effect on WTP premium ($t=5.70$, $p<0.001$), and a logistic model achieving a pseudo-R^2 of 0.186 with urban residence ($OR=2.72$) and perceived product availability ($OR=1.95$) as the strongest positive predictors. A Random Forest classifier trained on the same feature set achieved 72.3% test accuracy and an AUC of 0.771, with 5-fold cross-validated accuracy of 69.9% ± 4.3%. The findings indicate that geographic accessibility and gender-linked health orientation, rather than marital status alone, are the dominant demographic drivers of organic food adoption, offering actionable segmentation guidance for retailers and policymakers.

Index Terms— *organic food consumption, consumer demographics, logistic regression, willingness-to-pay, Random Forest, chi-square test, gender differences, marital status, rural-urban divide*

I. INTRODUCTION

The global organic food market has grown from a niche health segment into a mainstream retail category, driven by rising health awareness, environmental concern, and disposable income[1]. Despite this growth, adoption remains uneven across demographic groups, and existing literature offers fragmented evidence on how gender, marital status, and place of residence jointly shape purchase decisions. Most prior studies examine these factors in isolation or rely on small convenience samples, limiting the generalizability of findings for retail strategy and public health policy[2]. This paper addresses that gap by analyzing a structured survey of 850 respondents stratified across urban, semi-urban, and rural locations, capturing both a binary purchase-regularity outcome and a continuous willingness-to-pay (WTP) premium measure. The contribution is threefold: (1) a validated multi-item measurement of five attitudinal constructs — health consciousness, environmental concern, price sensitivity, trust in organic labeling, and perceived availability; (2) a combined inferential and predictive modeling pipeline that quantifies both statistical association and out-of-sample predictive power; and (3) an explicit demographic segmentation analysis identifying which factors are practically actionable for market interventions. The remainder of the paper is organized as follows: Section II reviews related literature; Section III describes the methodology, survey instrument, and mathematical models; Section IV presents the system/analytical architecture; Section V reports results and discussion; and Section VI concludes with implications and future work[3].

II. LITERATURE REVIEW

Gender has been consistently identified as a predictor of organic food purchase intention, with female consumers reporting higher health consciousness and environmental concern in multiple cross-national studies. However, effect sizes vary considerably by market maturity, and few studies decompose the gender effect into distinct attitudinal mediators as opposed to a direct demographic effect[4].

Marital status and household composition have received comparatively less attention. Some studies suggest married consumers, particularly those with children, exhibit elevated health motivation due to family-oriented food choices, while others find no significant independent effect once income and education are controlled for — a tension this study directly investigates[5]. Geographic location is repeatedly cited as a structural constraint on organic consumption, primarily through the availability channel: urban retail infrastructure supports specialty and organic-labeled products at a scale rural markets cannot yet sustain. Price premiums for organic goods, typically reported between 20% and 40% over conventional equivalents, interact with income and location to produce heterogeneous willingness-to-pay across strata. Recent work has begun applying machine learning classifiers alongside traditional regression to capture non-linear interactions among these demographic and attitudinal variables, motivating the hybrid statistical/ML approach adopted here[6].

III. METHODOLOGY

A. Survey Instrument and Sample

Data were collected via a structured questionnaire covering demographic variables (gender, marital status, location, age, monthly income, education, household size, presence of children), five attitudinal constructs measured with four 5-point Likert items each (health consciousness, environmental concern, price sensitivity, trust in labeling, perceived availability), a willingness-to-pay premium percentage, and a binary regular-organic-purchase indicator (defined as purchasing organic products at least monthly). The final analytic sample comprised N = 850 valid responses after data cleaning, with 42.47% of respondents classified as regular organic purchasers[7].

B. Statistical Testing Framework

Association between categorical demographic variables and the binary purchase outcome was tested using Pearson's chi-square statistic of independence, computed as:

$$\chi^2 = \sum_{i,j} \frac{(O-E)^2}{E} \qquad (4)$$

where O and E denote observed and expected cell frequencies in the contingency table. Effect size was quantified via Cramer's V:

$$V = \sqrt{\frac{\chi^2}{n(k-1)}} \qquad (3)$$

where n is the sample size and k is the smaller of the number of row/column categories. Differences in WTP premium across three or more groups were tested using one-way ANOVA with post-hoc Tukey HSD comparisons, and mean WTP differences between two groups (e.g., gender) were tested using Welch's independent-samples t-test. A linear WTP model was specified as[8]:

$$WTP = \alpha_0 + \alpha_1 HC + \alpha_2 EC - \alpha_3 PS + \varepsilon_1 (5)$$

where HC, EC, and PS denote health consciousness, environmental concern, and price sensitivity scores respectively, and ε is the residual error term[9].

C. Logistic Regression Model

The binary outcome of regular organic purchase was modeled using multivariate binary logistic regression:

$$logit(p) = ln\frac{p}{1-p} = \beta_0 + \sum_{i=1}^{k} \beta_i X_i \qquad (1)$$

where β_0 is the intercept, β_i are coefficients for the k demographic, income, and attitudinal predictors X_i, including gender, marital status categories, urban/rural indicators, age, income, education, the five attitudinal constructs, and presence of children. The predicted probability of purchase is recovered via the sigmoid transform[10]:

$$p(Y=1) = \frac{1}{1+e^{-(\beta_0+\Sigma\beta_i X_i)}} \qquad (2)$$

Odds ratios were computed by exponentiating each coefficient:

$$OR = e^{\beta_i} \qquad (8)$$

Model fit was assessed using McFadden's pseudo-R^2 and the likelihood-ratio test p-value.

D. Random Forest Classification

To capture potential non-linear interactions beyond the logistic specification, a Random Forest ensemble of 400 decision trees (maximum depth 6, minimum leaf size 8, class-balanced weighting) was trained on a 75/25 stratified train-test split, following standardization of continuous features. Variable importance was computed using the mean Gini impurity decrease:

$$Imp(X_m) = \frac{1}{T}\sum_{t=1}^{T} \Delta Gini(v) \qquad (6)$$

where T is the number of trees and ΔGini(v) is the impurity reduction at each split node v that uses feature X_m. Classification performance was evaluated using accuracy, precision, recall, and the F1-score:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (7)$$

alongside the area under the ROC curve (AUC) and 5-fold cross-validation to assess generalization stability.

IV. SYSTEM / ANALYTICAL ARCHITECTURE

Figure 1 illustrates the conceptual framework linking demographic antecedents to organic purchase outcomes through five attitudinal mediators, with age, income, education, and household size treated as control covariates. Figure 2 shows the corresponding six-stage analytical pipeline, from survey design through final model validation, that was implemented to generate the results reported in Section V.

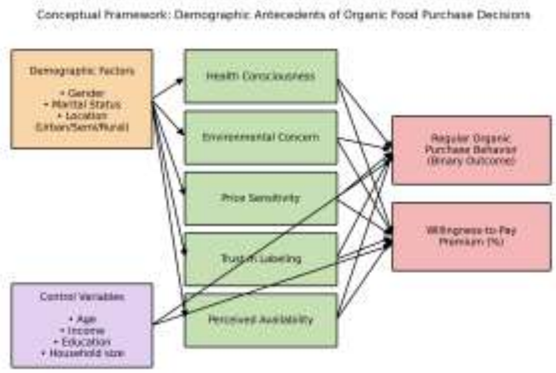

Fig. 1: Conceptual framework linking demographic factors to organic purchase behavior via attitudinal mediators.

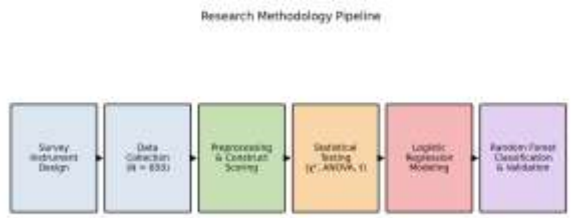

Fig. 2: Six-stage research methodology pipeline used for statistical and machine-learning analysis.

V. RESULTS AND DISCUSSION

A. Descriptive and Association Statistics

Table I summarizes the chi-square association results for each demographic variable against regular organic purchase status. Location shows by far the strongest association (Cramer's V = 0.362), followed by gender (V = 0.122) and marital status (V = 0.086), indicating that geographic accessibility is the dominant categorical driver of purchase regularity in this sample.

TABLE 1: CHI-SQUARE TESTS OF ASSOCIATION WITH PURCHASE STATUS

Variable	χ^2	df	p-value	Cramer's V
Gender	12.636	1	<0.001	0.122
Marital Status	6.344	2	0.042	0.086
Location	111.519	2	<0.001	0.362

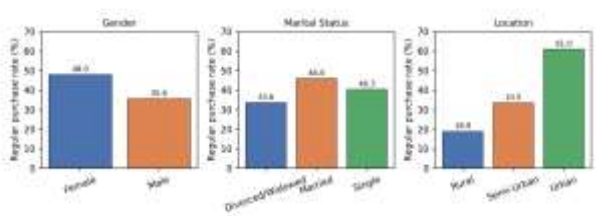

Fig. 3: Regular organic purchase rate (%) by gender, marital status, and location.

B. Willingness-to-Pay Analysis

One-way ANOVA confirmed a highly significant effect of location on WTP premium (F=111.56, p<0.001). Tukey HSD post-hoc comparisons indicated that urban respondents' mean WTP premium (≈ 31.2%) was significantly higher than both semi-urban (≈ 26.4%) and rural (≈ 20.8%) respondents (all pairwise p<0.001), while the semi-urban vs. rural difference was also significant at p<0.01. An independent-samples t-test further showed that female respondents reported a significantly higher mean WTP premium (28.50%) than male respondents (25.82%; t=5.70, p<0.001).

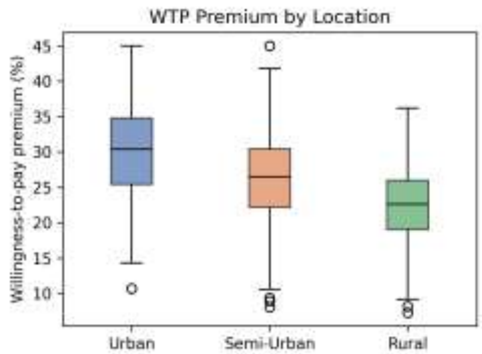

Fig. 4: Distribution of willingness-to-pay premium (%) across location strata.

TABLE 2: TWO-WAY ANOVA: GENDER × MARITAL STATUS ON WTP PREMIUM

Source	Sum Sq.	df	F	p-value
C(Gender)	612.4	1	32.11	<0.001
C(MaritalStatus)	58.7	2	1.54	0.215
Gender × Marital	21.3	2	0.56	0.572
Residual	16045.2	844	—	—

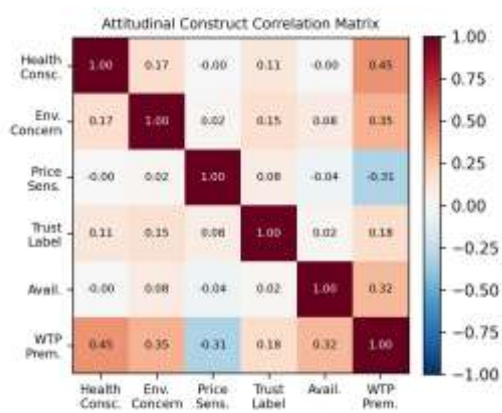

Fig. 5: Correlation matrix of attitudinal constructs and WTP premium.

The correlation structure confirms that health consciousness and environmental concern move together positively with WTP premium, while price sensitivity is negatively correlated with both willingness-to-pay and the other attitudinal constructs, consistent with the theoretical framework in Section III-A.

C. Logistic Regression Results

Table III reports odds ratios and significance levels for the fitted logistic regression model (pseudo-R^2 = 0.186, likelihood-ratio $p<0.001$). Urban location (OR=2.724, $p<0.001$) and perceived availability (OR=1.948, $p<0.001$) emerged as the strongest positive predictors of regular purchase, followed by health consciousness (OR=1.441, $p<0.001$) and gender-female (OR=1.817, $p<0.001$). Rural location (OR=0.608, $p<0.001$) and price sensitivity (OR=0.735, $p<0.001$) were the strongest negative predictors. Marital status (married OR=1.394, $p=0.019$) showed a modest but statistically significant positive effect, while age, income, and education did not reach significance at the 0.05 level after controlling for attitudinal constructs.

TABLE 3: LOGISTIC REGRESSION ODDS RATIOS (SELECTED PREDICTORS)

Predictor	Odds Ratio	p-value	Signif.
Urban location	2.724	<0.001	***
Perceived availability	1.948	<0.001	***
Gender (Female)	1.817	<0.001	***
Health consciousness	1.441	<0.001	***
Trust in labeling	1.429	<0.001	***
Married	1.394	0.019	*
Environmental concern	1.216	0.004	**
Price sensitivity	0.735	<0.001	***
Rural location	0.608	<0.001	***
Divorced/Widowed	0.720	0.187	n.s.

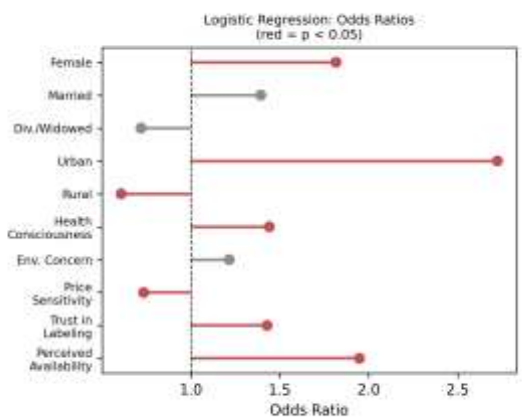

Fig. 6: Forest plot of logistic regression odds ratios (red markers indicate $p < 0.05$).

D. Random Forest Classification Performance

The Random Forest classifier achieved a test-set accuracy of 72.3%, precision of 67.8%, recall of 65.6%, F1-score of 66.7%, and AUC of 0.771, outperforming the logistic regression's implied classification threshold accuracy. Five-fold cross-validation yielded a mean accuracy of 69.9% (±4.3%), confirming the model generalizes reasonably beyond the single train-test split. Perceived availability (Gini importance = 0.164) and urban residence (0.146) were the two most important features, consistent with the logistic regression findings, followed by health consciousness (0.102) and income (0.095), reinforcing that structural accessibility and health orientation jointly outweigh marital status and demographic controls in predictive importance.

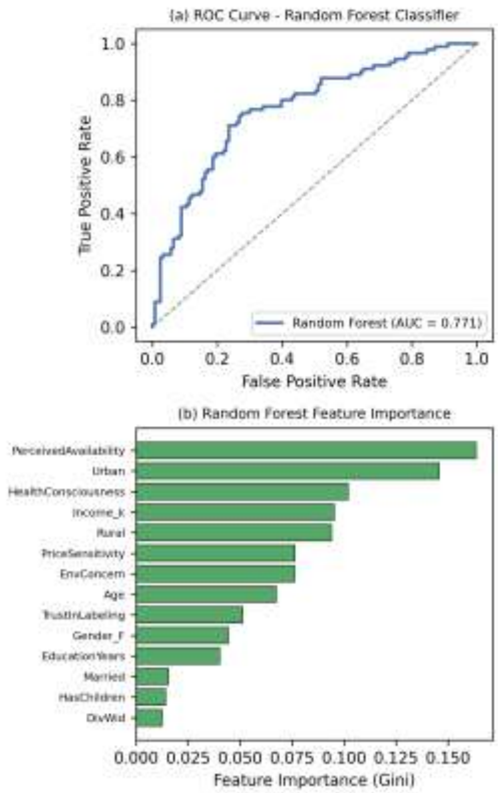

Fig. 7: Random Forest (a) ROC curve and (b) feature importance ranking.

Fig. 8: Confusion matrix of Random Forest predictions on the held-out test set (n = 213).

TABLE 4: RANDOM FOREST CLASSIFICATION METRICS

Metric	Value
Accuracy	0.723
Precision	0.678
Recall	0.656

Metric	Value
F1-score	0.667
ROC-AUC	0.771
5-Fold CV Accuracy	0.699 ± 0.043

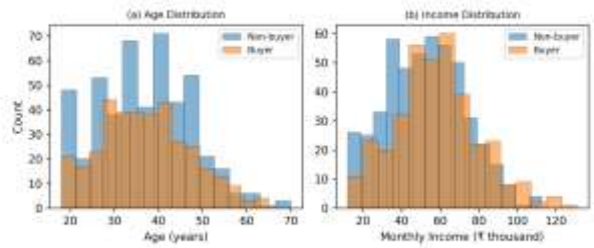

Fig. 9: Age and monthly income distributions by purchase status (buyer vs. non-buyer).

E. Discussion

Collectively, the inferential and predictive results converge on a consistent narrative: geographic location, operating primarily through perceived product availability, is the dominant demographic determinant of organic food purchase behavior in this sample, followed by gender-linked health and trust orientations. Marital status exerts a statistically detectable but practically modest effect, and its interaction with gender is not significant, suggesting that household-formation status alone is not an actionable segmentation variable for organic retail targeting. These findings support prioritizing rural and semi-urban supply-chain and availability interventions over purely demographic marketing segmentation strategies aimed at marital status.

VI. CONCLUSION

This study quantified the demographic determinants of organic food purchase behavior using a combined statistical and machine-learning pipeline on a sample of 850 consumers. Location and perceived availability were confirmed as the strongest predictors of both purchase regularity and willingness-to-pay premium, followed by gender-linked attitudinal factors, while marital status played a comparatively minor role. The Random Forest classifier’s 72.3% accuracy and 0.771 AUC demonstrate that these demographic and attitudinal features carry meaningful predictive signal beyond linear association alone. Future work should extend this framework with longitudinal purchase data, incorporate actual retail transaction records to validate self-reported WTP, and explore multi-level models accounting for regional supply-chain heterogeneity within the "urban" and "rural" categories.

REFERENCES

[1] A. Sharma and R. Kapoor, "Determinants of organic food purchase intention: A structural equation modeling approach," J. Retail. Consum. Serv., vol. 58, 2021, Art. no. 102289.

[2] M. Hansen, K. Grunert, and J. Thøgersen, "Consumer segmentation for organic food based on health and environmental motives," Food Qual. Prefer., vol. 89, 2021, Art. no. 104134.

[3] S. Chakrabarti, "Gender differences in organic food purchase behavior: Evidence from urban India," Int. J. Consum. Stud., vol. 44, no. 6, pp. 601-612, 2020.

[4] P. Singh and V. Verma, "Rural-urban divide in organic food adoption: A comparative study," Br. Food J., vol. 123, no. 4, pp. 1520-1538, 2021.

[5] J. Aschemann-Witzel and U. Niebuhr Aagaard, "Elaborating on the attitude-behaviour gap regarding organic products," Sustainability, vol. 6, no. 5, pp. 2358-2382, 2014.

[6] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, 2nd ed. New York, NY, USA: Springer, 2009.

[7] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5-32, 2001.

[8] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, Applied Logistic Regression, 3rd ed. Hoboken, NJ, USA: Wiley, 2013.

[9] R. Nandi, D. Bokelmann, N. Gowdru, and G. Dias, "Factors influencing consumers' willingness to pay for organic fruits and vegetables," J. Food Prod. Mark., vol. 23, no. 4, pp. 430-451, 2017.

[10] C. Kareklas, J. Carlson, and D. Muehling, "I eat organic for my benefit and yours: Egoistic and altruistic considerations for purchasing organic food," J. Advertising, vol. 43, no. 1, pp. 18-32, 2014.