

## Profiling U.S. Milk Consumers: A Machine Learning Assisted Audience Segmentation

Karissa Palmer<sup>1</sup>

[karissapalmer@tamu.edu](mailto:karissapalmer@tamu.edu)

Shuai Ma<sup>1</sup>

[shuai.ma@ag.tamu.edu](mailto:shuai.ma@ag.tamu.edu)

Zhihong Xu<sup>1</sup>

[zhihong.xu@ag.tamu.edu](mailto:zhihong.xu@ag.tamu.edu)

Holli Leggette<sup>1</sup>

[Holli.Leggette@ag.tamu.edu](mailto:Holli.Leggette@ag.tamu.edu)

Gary Wingenbach<sup>1</sup>

[Gary.Wingenbach@ag.tamu.edu](mailto:Gary.Wingenbach@ag.tamu.edu)

Oral Capps<sup>2</sup>

[oral.capps@ag.tamu.edu](mailto:oral.capps@ag.tamu.edu)

<sup>1</sup> **Texas A&M University**

Department of Agricultural Leadership, Education, and Communications

2116 TAMU

College Station, TX 77843-2116

<sup>2</sup> **Texas A&M University**

Department of Agricultural Economics

2116 TAMU

College Station, TX 77843-2116

**Funding Acknowledgment:** This material is based upon work supported by the Texas A&M AgriLife Institute for Advancing Health Through Agriculture (IHA) and the U.S. Department of Agriculture, Agricultural Research Service, under Agreement No. 58-3091-1-018. Any opinions, findings, conclusion, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the U.S. Department of Agriculture.

## Profiling U.S. Milk Consumers: A Machine Learning Assisted Audience Segmentation

### Introduction

The new dietary guidelines from the United States Department of Agriculture prioritize the consumption of whole, nutrient-dense foods (U.S. Department of Health and Human Services & U.S. Department of Agriculture, 2026). Dairy and plant-based milk offer differing nutrient profiles, and long-term trends show a decline in dairy consumption and an increase in plant-based beverages (Wolf et al., 2020). Consumers of plant-based milk choose alternative milk products for various reasons, including dietary preferences, taste, religious considerations, environmental concerns, or milk protein allergies/intolerances (Chalupa-Krebzdak et al., 2018). Although nutritional differences between dairy milk and plant-based milk may influence purchasing decisions (McCarthy et al., 2017), consumption and market for such foods largely depend on consumers' health consciousness and knowledge (Hoque et al., 2018). Milk consumers who are more environmentally conscious and concerned about animal welfare tend to pay closer attention to details, viewing their food as part of a healthy lifestyle (Marchini et al., 2021). Therefore, we identified milk consumer profiles to enhance targeted communication and marketing strategies based on demographic and psychographic characteristics.

### Methods

We used a cross-sectional survey to collect data from U.S. consumers. The survey, which was part of a larger study, had several sections we used to collect participants' demographic and psychographic information, including attitudes, milk-purchasing intentions, and milk preferences. We collected data through an online sample purchased via Qualtrics sampling services, obtaining 2,002 responses. To analyze the data, we employed unsupervised machine learning techniques, specifically K-means clustering in Python, using data from 2,002 consumers to segment participants into distinct groups based on selected features, including attitude, purchase intention, knowledge, belief, norm, benefit, risk, trust, health consciousness, and environmental concern. Clustering was used to identify naturally occurring consumer segments based on psychographic characteristics, allowing comparison of distinct consumer profiles. To determine the optimal number of clusters, we used the Silhouette Score and the Calinski–Harabasz Index as internal validation metrics to evaluate clustering quality. The Silhouette Score assesses how well observations fit within their assigned clusters by balancing within-cluster cohesion and between-cluster separation, whereas the Calinski–Harabasz Index evaluates overall cluster compactness and separation, with higher values indicating a clearer cluster structure. After identifying distinct clusters, each cluster was assigned a group label, and independent sample tests were used to confirm significant distinctions between clusters. Descriptive statistics were computed for demographic information and psychographic variables across the identified groups.

### Results

The Silhouette Score peaked at  $k = 2$ , indicating the strongest overall alignment between data points and their assigned clusters. The Calinski–Harabasz Index also showed a notable local maximum at  $k = 2$ . We selected  $k = 2$  as the optimal clustering solution based on these two metrics. An independent-samples t-test comparing the variables confirmed significant differences between the high-purchase intention group and the low-purchase intention group. All p-values between the two clusters were significant for *attitude, intention, knowledge, belief, norm, benefit,*

*risk, trust, health consciousness, and environmental concern.* Descriptive statistics across the two clusters reveal two distinct participant profiles, each with unique demographic compositions and psychographic attributes (see Figure 1).

**Figure 1.** *Milk Consumer Clusters*



Group 0 (high attitude and purchase intention group) consists of 40.93% ( $n = 467$ ) dairy milk consumers and 59.07% ( $n = 674$ ) plant-based consumers, indicating a dominant plant-based milk consumption pattern. They are younger (53.64% are aged 25–44), and most (66.70%) have children or dependents under 18 living at home, and a higher proportion of the population has incomes over 75,000. On the contrary, Group 1 (low attitude and purchase intention group) consists of 61.56% ( $n = 530$ ) dairy milk consumers and 38.44% ( $n = 331$ ) plant-based consumers, indicating a dominant dairy milk consumption pattern. The group is older (38.44% are aged 65 and over), and ~68% do not have children living at home. Many have high school and some college level education with an income of \$10,000 to \$74,999.

### Discussion and Conclusions

The cluster analysis reveals two distinct consumer segments with meaningful demographic and behavioral differences. Group 0, characterized by higher attitudes and purchase intentions, skews younger, more affluent, and likely to have children at home. These characteristics may contribute to their stronger adoption of plant-based milk. Their dominant plant-based consumption pattern suggests younger, higher-income families may be driving growth in the alternative-milk market. Marketing strategies for plant-based milk may be most effective when targeted at this population. In contrast, Group 1 represents an older, lower-intention segment that remains more loyal to dairy milk, with fewer at-home dependents and more moderate income and education levels. These findings imply that dairy milk brands may benefit from reinforcing familiarity and tradition among older demographics. The clear segmentation provides opportunities for tailored messaging, product positioning, and educational efforts to address differing motivations and purchasing barriers across populations.

### References

- Chalupa-Krebszdek, S., Long, C. J., & Bohrer, B. M. (2018). Nutrient density and nutritional value of milk and plant-based milk alternatives. *International Dairy Journal*, 87, 84–92. <https://doi.org/10.1016/j.idairyj.2018.07.018>
- Hoque, M. Z., Alam, M. N., & Nahid, K. A. (2018). Health consciousness and its effect on perceived knowledge, and belief in the purchase intent of liquid milk: Consumer insights from an emerging market. *Foods*, 7(9), 150. <https://doi.org/10.3390/foods7090150>
- Marchini, A., Riganelli, C., Diotallevi, F., & Polenzani, B. (2021). Label information and consumer behaviour: evidence on drinking milk sector. *Agricultural and Food Economics*, 9(1), 8. <https://doi.org/10.1186/s40100-021-00177-5>
- McCarthy, K. S., Parker, M., Ameerally, A., Drake, S. L., & Drake, M. A. (2017). Drivers of choice for fluid milk versus plant-based alternatives: What are consumer perceptions of fluid milk?. *Journal of Dairy Science*, 100(8), 6125–6138. <https://doi.org/10.3168/jds.2016-12519>
- U.S. Department of Health and Human Services & U.S. Department of Agriculture. (2026). *Dietary Guidelines for Americans, 2025–2030* (10th ed.). <https://cdn.realfood.gov/DGA.pdf>
- Wolf, C. A., Malone, T., & McFadden, B. R. (2020). Beverage milk consumption patterns in the United States: Who is substituting from dairy to plant-based beverages?. *Journal of Dairy Science*, 103(12), 11209–11217. <https://doi.org/10.3168/jds.2020-18741>