

# Enhancing Social Media Impact: Leveraging User Demographics to Understand Text Tonality Preferences

Elizabeth Soto<sup>1</sup>, Dr. Marcelo Sztainberg<sup>1</sup>, Dr. Nancy Wrinkle<sup>1</sup>, Dr. Benito Mendoza<sup>2</sup> (Faculty Advisors)

Northeastern Illinois University, Chicago, IL 60625<sup>1</sup>

New York City College of Technology, Brooklyn, NY 11201<sup>2</sup>



## Abstract

In the United States, the persistent challenge of food insecurity casts a shadow over many households, affecting 22 percent of Hispanic households. Overcoming language barriers and accessing essential food resources remains challenging for many in this demographic group. As social media platforms become primary sources of information about nearby food pantries, creating culturally resonant and engaging content becomes crucial for food security organizations. The goal of this study is to enhance the impact of social media content, particularly for food safety security organizations, by understanding the relationship between user demographics and text tonality preferences. We began by developing a two section survey. The first section of the survey collects information about participants' backgrounds, while the second part elicited preferences of text tonality. We collected internal tentative data and trained a hierarchical clustering algorithm. Our analysis identified three distinct clusters based on text tonality preferences. Within these clusters, notable preferences emerged. Hispanic female students predominantly favored empathetic posts, while non-binary participants showed a strong preference for persuasive posts in both English and Spanish. Interestingly, male Hispanic participants exhibited no significant preference for tonalities in posts but consistently favored English-language content, and many of them were employed part-time or full-time, unlike the other clusters that included students. The research team's qualifications, spanning user engagement, machine learning, data analysis, and linguistics, ensure a multidisciplinary and comprehensive approach to the research, enhancing its intellectual merit. This study holds the potential to benefit society because by studying the correlation between user engagement and tonal preferences, food safety security organizations can create content that effectively engages the Hispanic audience, which in turn can help apply these findings and reach those in the Hispanic community who struggle from food insecurity.

## Introduction/Background

In the United States, the persistent challenge of food insecurity casts a shadow over 22 percent of Hispanic households. Factors such as a lack of knowledge about sustainable food resources and language barriers often hinder Latino households from accessing private or community food assistance programs [1]. This troubling reality highlights the critical need for food security organizations to create social media posts that not only generate greater engagement but also effectively resonate with the Hispanic audience.

## Hypothesis/Intent

This study aims to analyze the correlation between user demographics and preferred text tonalities using the hierarchical clustering algorithm. The algorithm will consider data gathered from a survey, including age, ethnicity, educational background, employment status, and text tonality preferences. The survey will collect preliminary data collected from personnel who funded the study. The hypothesis is that the generated clusters will reveal significant patterns among different demographic groups.

## Materials/Methods

- **Hierarchical Clustering**: A method of cluster analysis in data mining that creates a hierarchical representation of the clusters in a dataset. The method starts by treating each data point as a separate cluster and then iteratively combines the closest clusters until a stopping criterion is reached. [5]
- **Cophenetic Correlation**: A measure of how faithfully a dendrogram preserves the pairwise distances between the original unmodeled data points. [6]

## Methodology

1. Participants were presented with eight questions, each featuring a social media post from the Food Safety Security Organization.
2. For each question, participants were offered four randomly arranged text options, each representing a modification of the original post with different tonalities (empathetic, persuasive, simpler, and the original tone).
3. Collected preliminary data from the research group.
4. Prepared the dataset for analysis by performing data cleansing, encoding, and feature engineering.
5. Utilized the dataset to create a dendrogram, aiding in the determination of the optimal number of clusters.
6. Trained the Hierarchical clustering algorithm using both the dataset and the optimal number of clusters.
7. Evaluated the clustering results using the cophenetic score to measure the accuracy of the predictions.
8. Utilized visualizations to examine the correlation between user demographics and tone preferences within each cluster.

## Results

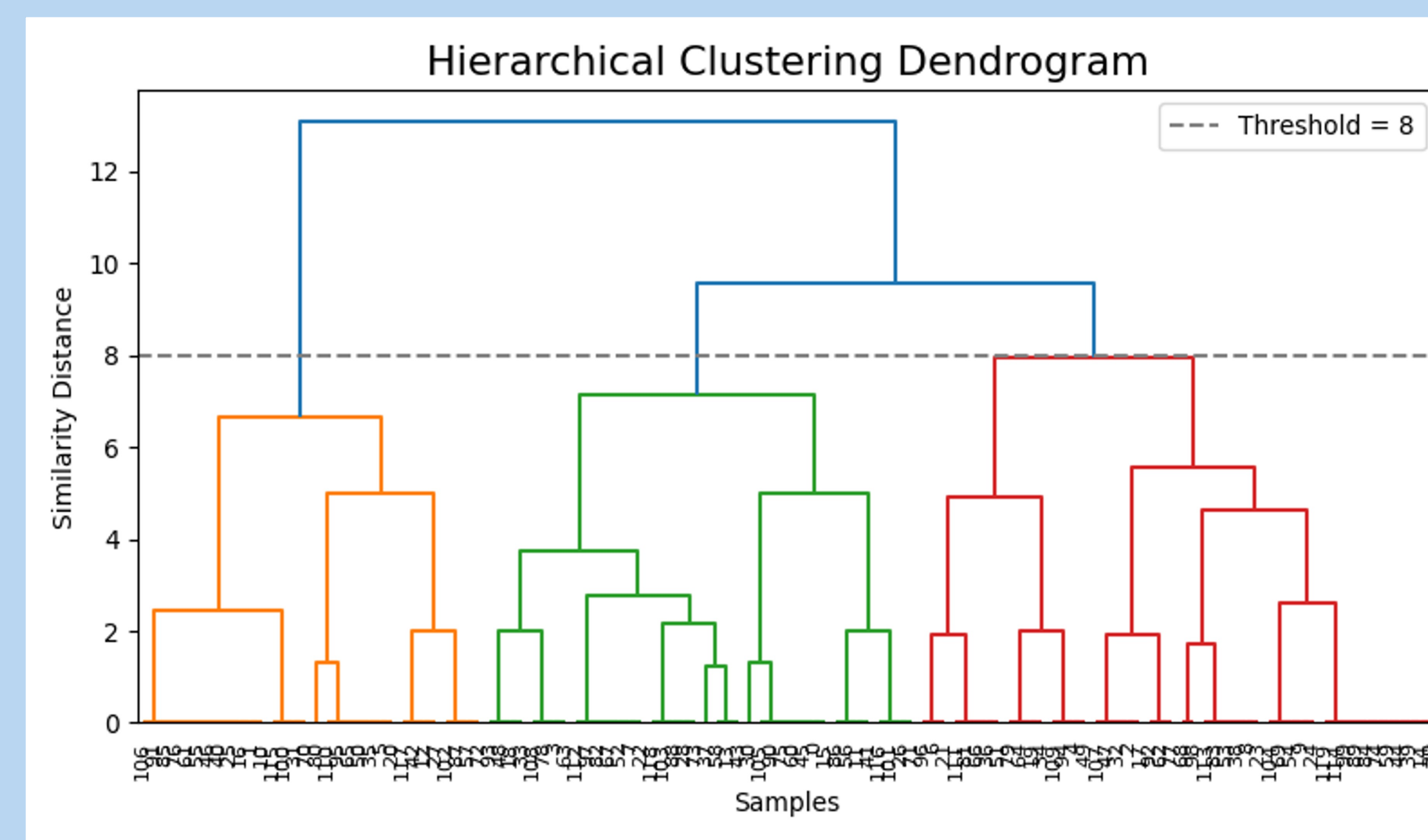


Figure 1.1 Hierarchical clustering dendrogram

- The Y-axis in figure 1.1 represents the dissimilarity or distance between data points or clusters. As you move up the Y-axis, the dissimilarity between clusters increases, which means that clusters become less similar as you go higher on the Y-axis
- The X-axis represents individual data points. To determine the number of clusters, we set a threshold and cut through the dendrogram at a height of eight. The number of clusters are equal to the count of vertical lines intersected by the threshold, which is three.

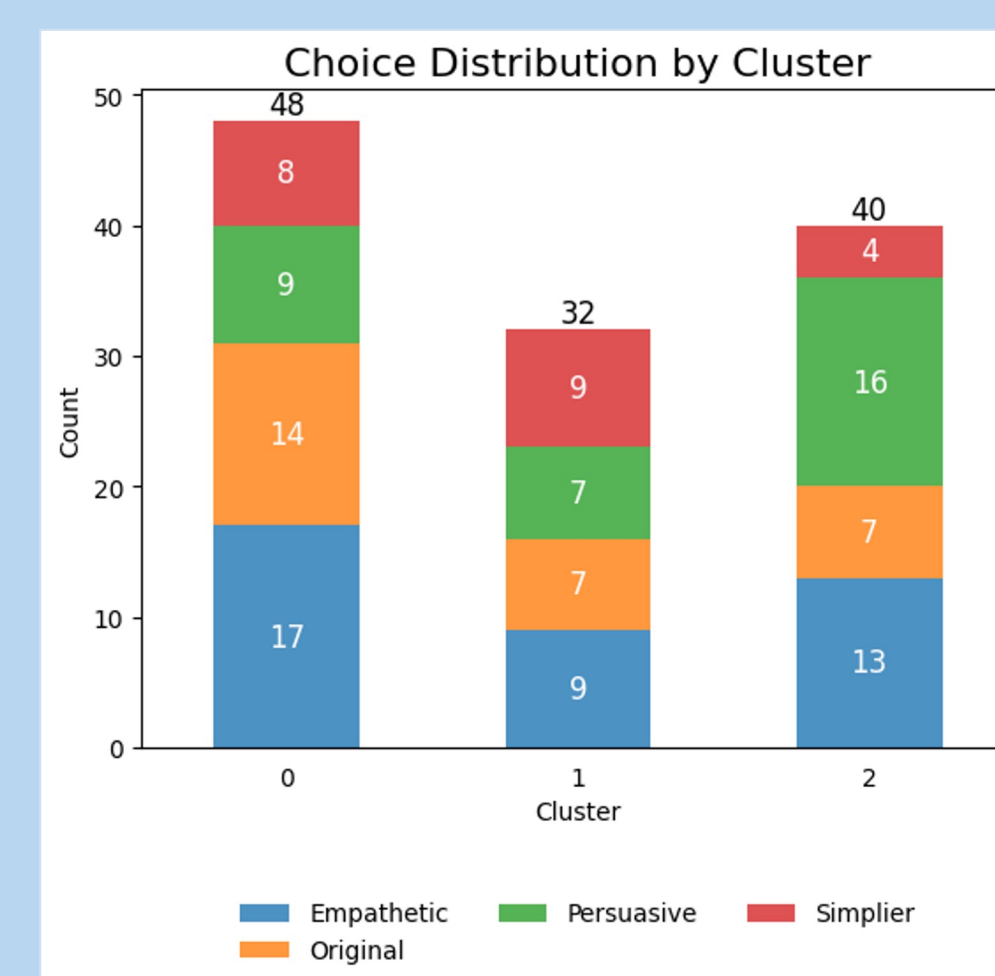


Figure 1.2 Tone preference cluster distribution

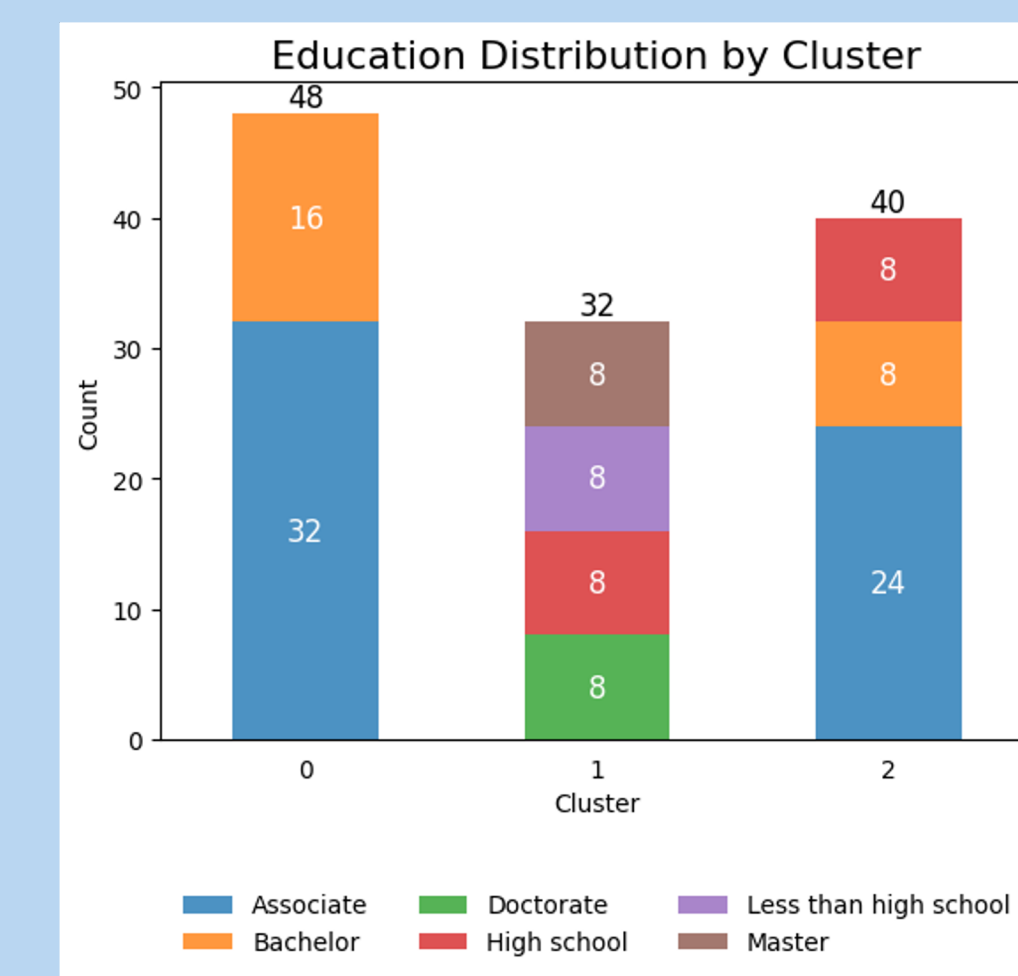


Figure 1.3 Education cluster distribution

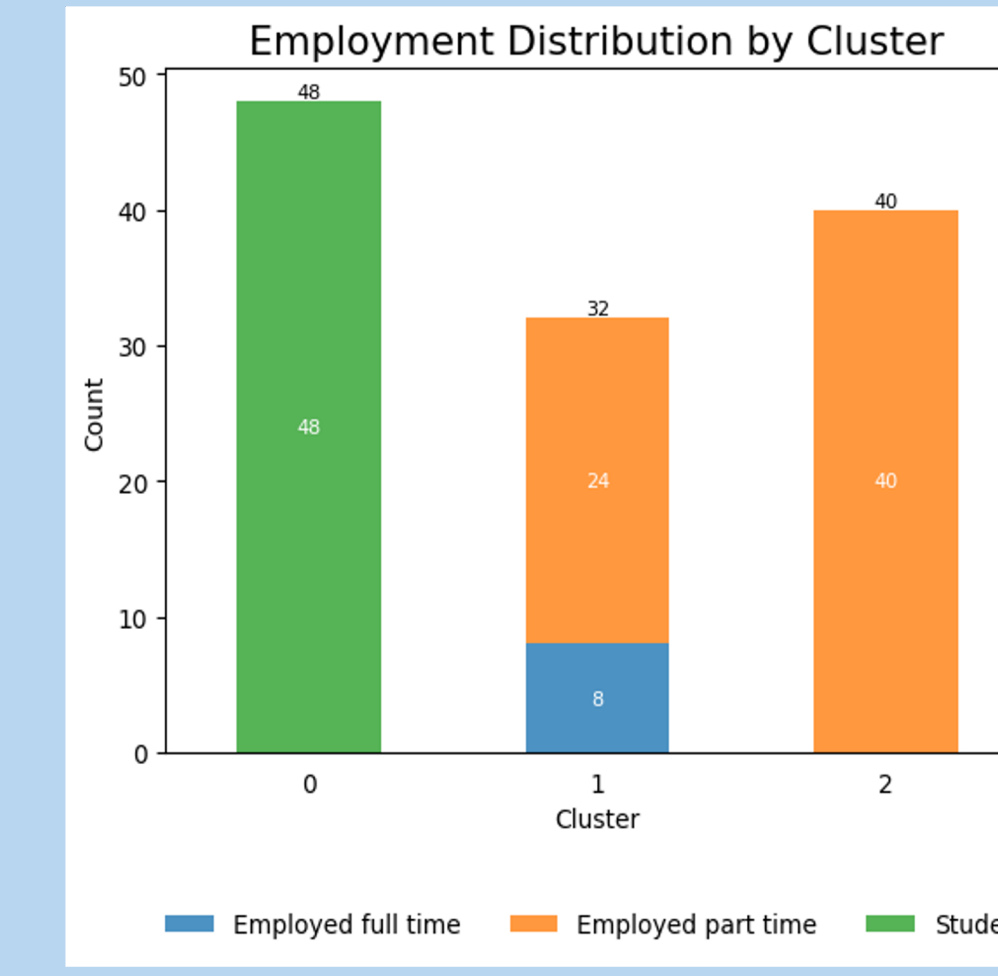


Figure 1.4 Employment cluster distribution

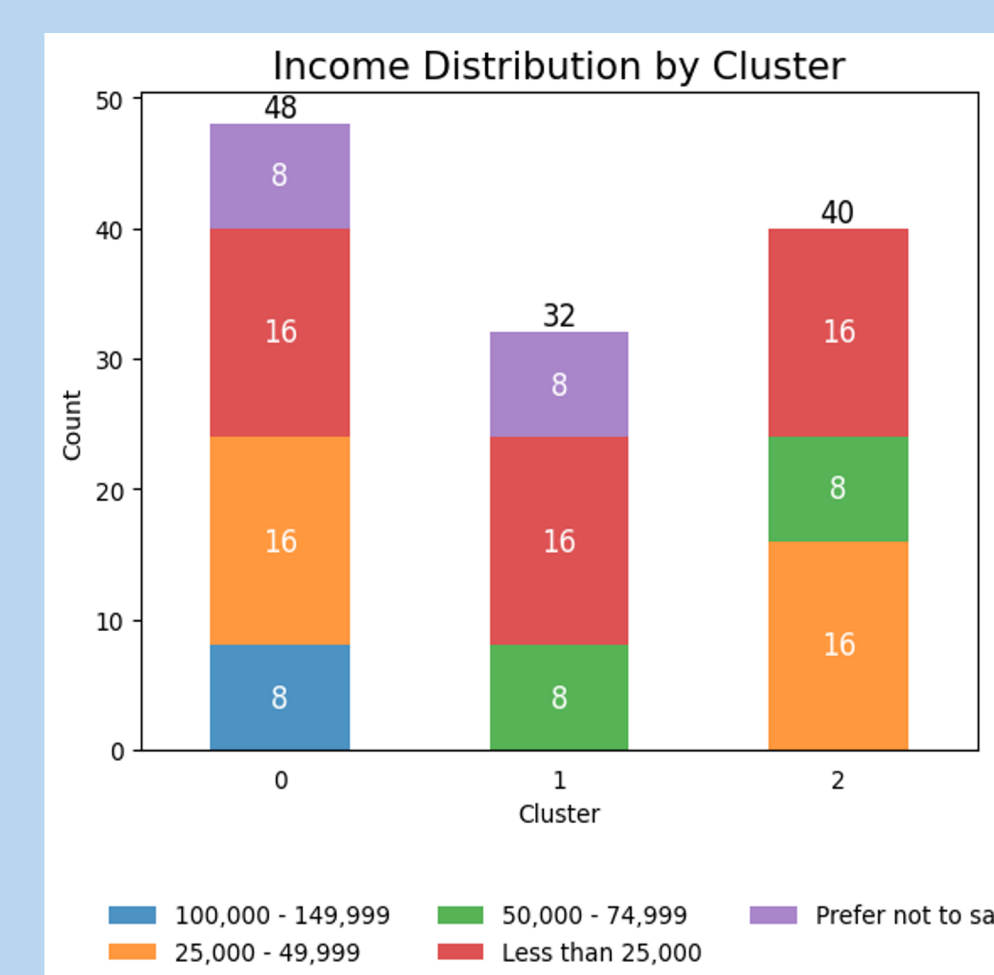


Figure 1.5 Income cluster distribution



Figure 1.6 Age cluster distribution

## Results

- The model revealed three distinct clusters based on participants' preferences.
- Hispanic female students predominantly favored empathetic posts
- Non-binary participants showed a strong preference for persuasive posts in both English and Spanish.
- Male Hispanic participants exhibited no significant preference for tonalities in posts but consistently favored English-language content, and many of them were employed part-time or full-time, unlike the other participants who were primarily students.
- We obtained a cophenetic score of 0.85, which suggest that the clustering result are an excellent representation of the original distances.

## Conclusion

Our study's implications for content personalization and audience targeting underscore its transformative potential. By understanding the correlation between tonality and engagement, we contribute to the development of impactful user engagement methods, with a focus on addressing food insecurity in the Hispanic community.

## Future Work

In the future, we aim to explore demographics further by examining finer attributes beyond ethnicity. We intend to concentrate on different combinations of demographic factors to create more detailed clusters. This approach will help us better understand how various demographic characteristics interact with tonality preferences in our survey participants.

## References

[https://docs.google.com/document/d/1T\\_3QWC5ikTvx3QIV77gZrHjV5s28\\_MA5qmYDHqY9F08/edit?usp=sharing](https://docs.google.com/document/d/1T_3QWC5ikTvx3QIV77gZrHjV5s28_MA5qmYDHqY9F08/edit?usp=sharing)

## Acknowledgements

This material is based upon work supported by the National Science Foundation under Grants numbers 2131291. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. I extend my sincere appreciation to Maya Papayan for generously providing the text samples that were instrumental in our analysis of user tonality preferences. Additionally, I am grateful to my mentors for their unwavering support and guidance throughout this project.